DOI:
10.1007/s10506-014-9157-3

Document Version
Peer reviewed version

Citation for published version (APA):
Keppens, J. (2014). On Modelling Non-Probabilistic Uncertainty In The Likelihood Ratio Approach To Evidential Reasoning. *Artificial Intelligence and Law*, *22*(3), 239-290. https://doi.org/10.1007/s10506-014-9157-3

Self-archived copy of a paper published in Artificial Intelligence and Law (2014) 22: 239–290.
The final publication is available at Springer via http://dx.doi.org/10.1007/s10506-014-9157-3

On Modelling Non-Probabilistic Uncertainty In The Likelihood Ratio Approach To Evidential Reasoning

Jeroen Keppens

Abstract When the likelihood ratio approach is employed for evidential reasoning in law, it is often necessary to employ subjective probabilities, which are probabilities derived from the opinions and judgement of a human (expert). At least three concerns arise from the use of subjective probabilities in legal applications. Firstly, human beliefs concerning probabilities can be vague, ambiguous and inaccurate. Secondly, the impact of this vagueness, ambiguity and inaccuracy on the outcome of a probabilistic analysis is not necessarily fully understood. Thirdly, the provenance of subjective probabilities and the associated potential sources of vagueness, ambiguity and inaccuracy tend to be poorly understood, making it difficult for the outcome of probabilistic reasoning to be explained and validated, which is crucial in legal applications. The former two concerns have been addressed by a wide body of research in AI. The latter, however, has received little attention. This paper presents a novel approach to employ argumentation to reason about probability distributions in probabilistic models. It introduces a range of argumentation schemes and corresponding sets of critical questions for the construction and validation of argument models that define sets of probability distributions. By means of an extended example, the paper demonstrates how the approach, argumentation schemes and critical questions can be employed for the development of models and their validation in legal applications of the likelihood ratio approach to evidential reasoning.

Keywords Evidential Reasoning, Bayesian Reasoning, Argumentation

1 Introduction

In AI and Law, approaches to evidential reasoning can be classified according to three categories denoting the type of methodology employed: argumentative, narrative and probabilistic. Argumentative evidential reasoning is concerned with decomposing the reasons to agree or disagree with information inferred from the available evidence

Jeroen Keppens, King's College London, Department of Informatics, Strand, London WC2R 2LS, UK,
E-mail: jeroen.keppens@kcl.ac.uk

into distinct arguments and to analyse their interrelationships [45]. Narrative evidential reasoning aims to produce a coherent story to explain the available evidence [7]. Probabilistic evidential reasoning seeks to assess the value of individual pieces of evidence in deciding which of a number of conflicting hypotheses to accept [11,36]. Increasingly, different evidential reasoning approaches are combined with a view to tackle more complex questions or to address particular weaknesses of individual approaches.

This article is concerned with the development of probabilistic models for evidential reasoning that aim to assess the value of evidence. This is challenging for a number of reasons. Firstly, the meaning of assessments of the value of evidence derived from a probabilistic model and the context within which it must be considered is often misunderstood [18]. Secondly, as models of complex problems, probabilistic models are abstractions of real-world scenarios. It can be difficult to ensure that all the qualitatively significant variables and relationships between these variables have been included.

Thirdly, probabilistic models compute joint probability distributions that assume that the probability distributions of certain variables are independent from one another under certain conditions [44]. It can be difficult to ensure that appropriate independence assumptions are made. Various authors have sought to tackle these concerns. For example, Fenton et. al. [25], Hepler et. al. [32] and Vlek et. al. [57] have identified and formalised key idioms employed in Bayesian models of evidential reasoning and developed methods to compose these into Bayesian models. Fenton et. al. and Hepler et. al. formalise these by means of composable fragments of Bayesian networks whereas Vlek et. al. employ narratives that are then converted into Bayesian networks. Keppens has presented an algorithm designed to extract arguments from probabilistic (Bayesian) models as a means to aid the validation of the structure of such models [35].

Fourthly, probabilistic models require probability distributions as input and some of these probability distributions are based on subjective human judgement, which may be flawed. Forensic statisticians employ sensitivity analysis to assess the effect of perturbations in such probabilities on the outcome of an analysis [2]. But while Bayesian inference mechanisms have been analysed and validated extensively [24], it is not unreasonable to question subjective probability values, the reasons for variability in such values and its magnitude. To facilitate such questioning by legal professionals, it is helpful to reduce subjective probability values and their variability to understandable propositions that can be attacked. This paper aims to accomplish this by incorporating argumentation as a means to reason about and validate subjective probability in a symbolic manner.

The approach introduced in this paper is based on a proposal by Druzdzel and van der Gaag to represent knowledge with regard to probability distributions by means of constraints [20]. Druzdzel and van der Gaag's motivation for such an approach is that it enables qualitative and quantitative information about probability distributions to be combined. In other words, these constraint based models can express probabilistic knowledge with varying degrees of precision. As such, this approach provides the same form of validation as a sensitivity analysis aiming to determine how the outcome of an analysis changes as probability values in the model vary [27], except that that variations in probability values are entered at the start and are assumed to stem from an expert's inherent uncertainty in providing those values. Neither Druzdzel and van der Gaag's constraint based approach nor sensitivity analysis provide a means to scrutinise the underlying reasons for choosing certain constraints or probability value ranges, however. As such, these approaches can compute the potential implications in correct probability estimations, but they cannot identify reasons why probability distribution estimates might be incorrect.

Models of argumentation have been used effectively as a means of scrutinising human reasoning. In particular, argumentation schemes (generic models of justifications for deriving certain conclusions based on a set of circumstances, observations and first principles) with corresponding sets of critical questions (schemes of attacks on the aforementioned justifications) have been shown to be particularly effective in the validation of evidential reasoning [8,52]. This paper shows how arguments supporting constraints over probability distributions can be defined and used to scrutinise probabilistic models of evidential reasoning. It introduces a set of argumentation schemes for defining such arguments and sets of associated critical questions to support their validation. The suitability of the set of defined argumentation schemes is demonstrated by means of a case study in which the probability distributions in a complex evidential reasoning model from the forensic statistics literature are redefined by means of argumentation models instantiated from the argumentation schemes.

2 Background

2.1 The Likelihood Ratio Approach to Compute the Value of Evidence

A likelihood ratio expresses how strongly evidence supports one hypaothesis over another, in a manner that is independent of prior beliefs in the hypotheses. Let h_1 and h_2 be two mutually exclusive and exhaustive hypotheses, both put forward as alternative explanations of a piece of observed evidence e. The likelihood of a hypothesis h_i is an assessment of that hypothesis's capacity to produce observation e [40]. In other words, the likelihood of h_i expresses how plausible it would be to encounter evidence e in a possible world where h_i holds true, irrespective of the plausibility or ones belief in h_i. The likelihood ratio $LR(e|h_1,h_2)$ is an assessment of the relative ability of the alternative hypotheses h_1 and h_2 to explain evidence e. In other words, $LR(e|h_1,h_2)$ determines how much better/weaker an explanation h_1 is for e compared to h_2.

The likelihood ratio is important to evidential reasoning in law because it represents an assessment of the value of evidence, from the perspective of a neutral observer who ought not to be affected by prior beliefs in h_1 or h_2. In other words, *taken by itself*, a likelihood ratio says nothing about the odds of hypotheses. Instead, the likelihood ratio describes how the odds of hypotheses change in response to the evidence. For this reason, this approach is favoured by many forensic scientists as the basis for presenting evidence in court [12].

It is common to express the likelihood of h_i as the probability of e given that h_i is presumed true: $P(e|h_i)$. Such a probability satisfies the aforementioned definition as (i) $P(e|h_i)$ measures the plausibility of e given h_i and (ii) $P(e|h_i)$ is independent of the prior probability $P(h_i)$ because h_i is set to true in order to compute $P(e|h_i)$. It follows that the likelihood ratio is [2]:

$$LR(e|h_1,h_2) = \frac{P(e|h_1)}{P(e|h_2)} \tag{1}$$

With this definition, the likelihood ratio expresses how the posterior odds of the hypotheses (in light of the evidence) are related to the prior odds of the hypotheses (before evidence was available). This follows from a simply rewriting of Bayes's theorem:

$$\frac{P(h_1|e)}{P(h_2|e)} = \frac{P(e|h_1)}{P(e|h_2)} \times \frac{P(h_1)}{P(h_2)}$$

or:

$$\text{Posterior Odds of Hypotheses} = LR \times \text{Prior Odds of Hypotheses} \tag{2}$$

Equation (2) shows how the likelihood ratio affects the odds of hypotheses.

In many evidential reasoning problems, probabilistic models to compute the numerator and denominator of (1) that are both precise and accurate, are not available. However, the likelihood ratio approach to compute the value of evidence need not be restricted to evidential reasoning problems where precise and accurate models are available [40]. Probabilistic models that account for the vagueness of the knowledge they are based on, allow for the likelihood ratio and the corresponding assessment of the value of evidence to be computed accurately, albeit less precisely [34]. But, problems may arise when this vagueness is not accounted for adequately and the model suggests a greater precision of analysis than the available model allows.

2.2 R v T & Argumentation Based Bayesian Evidential Reasoning

There have been a number of high profile appeal court cases where invalid applications of probabilistic evaluations of evidence in general and the likelihood ratio approach in particular have been the cause of miscarriages of justice. One recent and representative example is the judgement in *R v T*, a case heard in the High Court of Justice, Court of Appeal [4]. This case has gained notoriety in the evidential reasoning and forensic science communities because it imposes severe restrictions on the use of the likelihood ratio approach. A range of eminent researchers in this field have reported and discussed various flaws in the reasoning that lead to this conclusion [6,23,40,47,50,54]. However, for the purposes of this paper, I am interested in the specific appeal considered in the judgement in *R v T*, and the Court's reasons for quashing the original conviction.

At the original trial, T had been identified as the murderer of a victim on the basis of footwear marks (e) left at the scene of the crime. A forensic expert in evaluating footwear mark evidence had compared the footwear marks with footwear belonging to T. Following the likelihood ratio approach to evaluation of footwear mark evidence [12,22], the expert compared the likelihood that T's footwear made the footwear mark ($P(e|h_1)$) to the likelihood that some other shoe made the footwear mark ($P(e|h_2)$). In accordance with best practice among footwear experts in England and Wales, four factors were considered to compute these probabilities: sole pattern, shoe size, amount of wear and unique features, such as scratches and cuts caused by damage. The expert estimated that about 1 in 5 shoes would have a sole pattern consistent with the footwear mark recovered from the crime scene, about 1 in 10 shoes would leave the size 11 footwear mark recovered from the crime scene, about 1 in 2 shoes would exhibit wear consistent with that observed in the footwear mark and no shoes could be excluded on the basis of individual characteristics. On that basis, the likelihood ratio in this case was calculated to be:

$$LR = \frac{P(e|h_1)}{P(e|h_2)} = \frac{1}{\frac{1}{5} \times \frac{1}{10} \times \frac{1}{2} \times 1} = 100 \qquad (3)$$

On the basis of this analysis, and the reporting standard presented in [12], the expert reported:

> ”[...] there is at this stage a moderate degree of scientific evidence to support the view that the [footwear belonging to T] had made the footwear marks [...]” [4]

There are a number of concerns about this conclusion and the reasoning that led up to it, which are presented in the judgement. The first concern is that the conclusion of the expert can be interpreted as a statement about the probability that the footwear belonging to T had made the footwear marks or the probability that T had committed the crime. But, this is not the type of conclusion that can be derived from a likelihood ratio. Another way of verbalising the notion $LR = 100$ might be:

> ”The footwear mark evidence, taken by itself, provides a moderate degree of support for the view that the footwear mark belonging to T made the footwear marks at the crime scene, as opposed to the view that a random item of footwear made those marks.”

There are two important features of a likelihood ratio that need to be conveyed. On the one hand, the likelihood ratio compares two competing propositions or hypotheses. The nature of these propositions is important. In this case, the likelihood ratio compares one feature of the prosecution hypothesis, i.e. that a specific shoe made a mark, with a competing feature of the defence hypothesis, i.e. that the footwear mark was made by another shoe. On the other hand, as shown in (2), the likelihood ratio relates posterior odds to prior odds. It does not state whether a jury ought to believe one hypothesis over another. Instead, it states how such beliefs ought to be affected by the evidence as this is what an expert in evidence is expected to comment on. In this instance, without any further identifying information, it would be reasonable to believe that the prior odds of the prosecution hypothesis are very small relative to the defence hypothesis because the prosecution hypothesis argues that one individual pair of footwear made the marks whereas the defence hypothesis allows for a very large collection of shoes that could have made the marks. The likelihood ratio of (3) alters those odds by a factor 100 but this would not be enough to alter the favoured hypothesis in this scenario. The quote from the original trial, as reported in R v T does not appear to convey this clearly. Various other papers discuss ways in which a likelihood ratio analysis of forensic evidence could be reported better [26,40].

The second concern is that the likelihood ratio and the probabilities underpinning it might convey an inaccurate amount of precision. Various arguments are raised in R v T that variations in the proportion of patterns, sizes and wear encountered in different data sets might lead a forensic examiner to formulate different conclusions about the value of the evidence:

> ”The jury were given two important statistics based on the population as a whole:
> a) for pattern type of 0.25% or 1 in 400; this was given by the judge, [...]
> b) for size of 1 in 33; this was elicited in cross examination [...]’
> [...]
> If the figures used in the evidence for pattern type (1 in 400) and for size (1 in 33) are multiplied together [...], the result would be a likelihood ratio of 1 in 13,200 as opposed to the likelihood ratio of 1 in 50 used by [the forensic examiner], based on his more conservative approach. If the figure for wear [of 1 in 2] used by [the forensic examiner] was then applied [...], then the likelihood ratio would be 26,400 – in the verbal scale ”very strong support.” ” [4]

In other words, one unmoderated data set could be employed to conclude ”very strong support” for the hypothesis that footwear belonging to T made the footwear marks at the crime scene, as opposed to the hypothesis that a random item of footwear made those marks. At this stage, it is important to point out that, as stated in the above quote, the jury were given these low probabilities of $\frac{1}{400}$ and $\frac{1}{33}$ in support of the conclusion of ”moderate degree of scientific evidence to support the view that the [footwear belonging to T] had made the footwear marks”. However, this conclusion was supported by more conservative estimates and not by those low, precise point-probabilities. This is one of the reasons the Appeal Court concluded that ”The process by which evidence was adduced lacked transparency” [4]. Clearly, a jury cannot be relied on to interpret an evaluation based on a model where they are given inputs and outputs that are unrelated. This is one of the key reasons to quash the conviction.

The judgement goes on to claim: ”[...] it would be difficult to see how an opinion of footwear marks arrived at through the application of a formula could be described as ’logical’, or ’balanced’ or ’robust’, when the data are as uncertain as we have set out and could produce such different results” [4]. Thus, it is the judgement’s contention that, if there is a range of possible values for parts or all of the probabilities in a Bayesian model, then no reliable outcome can be produced with such a model. This argument is supported by various examples that show how different inputs based on different sources lead to very different outcomes. The assumption appears to be that a Bayesian model can only be relied on if there is one set of accurate and precise point probabilities. This is not correct. There is a rich body of literature on qualitative and semi-quantitative probabilistic reasoning that demonstrates that accurate probabilistic reasoning is possible with imprecise probabilities, provided that the vagueness of the knowledge on the basis of which imprecise probabilities are formulated is modelled accurately itself [13,21,42]. In testimony, the footwear expert revealed that the probabilities employed to determine that $LR = 100$ are conservative estimates. In other words, they are upper bounds. As such, a more accurate way of representing the probabilities for pattern type, size and wear would be by means of the intervals $[0, \frac{1}{5}]$, $[0, \frac{1}{10}]$ and $[0, \frac{1}{2}]$. This would support a range of likelihood ratios $LR \geq 100$, where the vagueness is due to lack of data. Such intervals reflect the very substantial amount of uncertainty and vagueness in this domain.

The third concern pertains to the use of ancillary evidence to determine the probabilities in the model. Direct evidence is evidence that informs assessments about the likelihood of hypotheses [52]. In the case under discussion, the footwear mark is direct evidence. Ancillary evidence is evidence that affects the ability of direct evidence to inform its assessment of likelihood [52]. In the case under discussion, the evidence used to estimate the probabilities for pattern type, size and wear is ancillary. Probabilistic models do not make a distinction between direct and ancillary evidence [45]. In fact, they tend to incorporate a significant amount of information derived from the ancillary evidence in the probabilities without referring to it any further. This can lead to a lack of transparency as has occurred in R v T, which is criticised in the judgement.

Specifically, the appeal court judgement reports a number of arguments against the feasibility of computing upper bounds on the probability that certain features are encountered in randomly selected items of footwear. For example, with regard to the likelihood of the footwear pattern encountered at the crime scene, the judgement notes that:

i) The brand of footwear of the suspect is counterfeit frequently and little data is available on the types and distribution of counterfeit footwear. It is argued that that makes it impossible to compute the probability of encountering shoes that possess similar features to those of T.
ii) The footwear pattern found at the crime scene is not only used by the major footwear manufacturer who produced T's footwear, but also by smaller footwear manufacturers. No data is available on the smaller footwear manufacturers. It is argued that that makes it impossible to compute the probability of encountering the sole pattern found at the crime scene.
iii) There are substantial local differences in the distribution of shoes. Local distributions of footwear depend substantially on the types of footwear sold by local shoe stores. It is argued that this makes it impossible to compute the probability of encountering footwear with a particular sole pattern in a given location, using only data on the national distribution of footwear.

The extent to which these concerns have been considered in reaching the eventual conclusion, as it is expressed above, or how the acceptance of any of these counterarguments would affect that conclusion remains unclear. Whereas a Bayesian model of evidential reasoning is able to address the previous two concerns, it provides no means of incorporating the reasons for accepting or rejecting significant features of conditional probability distributions, other than to implement their presumed implications numerically by adjusting the probability distributions. This is not satisfactory because, as the judgement argues, ”the process by which the evidence was adduced lacked transparency” [4].

This paper aims to address the latter concern by incorporating elements of argumentation into probabilistic models. As explained above, the transparency concern stems from a failure to validate the ancillary evidence that affects the probability values employed in a probabilistic model. The judgement in R v T does not show, in my opinion, that this failure is intrinsic to probabilistic models. Instead, it shows the difficulties of validating probabilistic models without adequately capturing ancillary evidence and reasoning based on ancillary evidence. Therefore, this paper seeks to address these difficulties by developing a means to incorporate and validate ancillary evidence and the way it affects probabilities into a probabilistic model.

3 Approach

3.1 Probabilistic Models

A variety of probabilistic modelling methods can be used to compute the likelihood ratio in the likelihood ratio approach to evidential reasoning. The work presented herein does not constrain which of these methods is employed. Throughout this paper the work will be illustrated by means of a Bayesian network (BN) designed to compute the likelihoods of prosecution and defence hypotheses. In complex evidential reasoning scenarios, BNs are the most widely used representation formalism to develop probabilistic models [29] because they allow joint probability distributions over a large number of variables to be specified and validated with relative easy. BNs are also of particular interest to the work presented herein, which combines argumentation with probabilistic modelling, because it has been shown that BNs can be used to model and reason with typical concerns that argumentation models of evidential reasoning deal with [25, 32].

3.1.1 Computing Likelihoods with Bayesian Networks

A BN provides a means to compute the likelihood of a hypothesis h with respect of a piece of evidence e as specified by the probability $P(e|h)$. It consists of a directed acyclic graph (DAG) and a set of probability distributions. Each node of the DAG corresponds to a variable that affects the value of $P(e|h)$. This variable may be discrete or continuous. The edges of the DAG specify relationships of probabilistic independence that limit the variable-value assignments that the probability distribution of each variable needs to be conditioned on. Specifically, to compute any conditional or joint probability distribution over variables of a BN, the probability distribution of each variable in the BN only needs to be conditioned on the possible combinations of value assignments of its immediate parent variables in the DAG.

The model of probabilistic independence specified by the DAG of a BN comes with a relatively straightforward intuitive meaning. This is perhaps most easily understood in the context of the most basic ways in which three variables X, S and Y can be related by a DAG: a serial connection of the form $X \rightarrow S \rightarrow Y$, a diverging connection of the form $X \leftarrow S \rightarrow Y$ and a converging connection of the form $X \rightarrow S \leftarrow Y$, such that there is no edge between X and Y in any of these three substructures.

In a serial connection $X \rightarrow S \rightarrow Y$, X and Y are probabilistically independent from one another provided the value of S is known. In a diverging connection $X \leftarrow S \rightarrow Y$, X and Y are also probabilistically independent from one another provided the value of S is known. In a converging connection $X \rightarrow S \leftarrow Y$, X and Y are probabilistically independent from one another if the values of S and all of its descendants are *not* known. In other words, X and Y are probabilistically *dependent* from one another, provided the value of S or one of its descendants is known. Each of these dependence or independence relationships is relatively easy to check and understand, making them easy to validate.

The advantage of using a BN to specify a probabilistic model of an evidential reasoning problem is that it facilitates the development of complex models. This is perhaps best illustrated by means of a practical example.

3.1.2 A Practical Application: Cross Transfer of Trace Evidence

This section presents a practical application of a BN to produce a model to compute the likelihood of a hypothesis. The model will be used throughout the remainder of this paper to illustrate the approach presented herein. It is a minor variation on the cross transfer of trace evidence model by Aitken, Taroni and Garbolino [3].

The model concerns a scenario involving a violent crime, where there is a victim, a suspect, a prosecution hypothesis that the suspect committed the crime and a defence hypothesis that the suspect did not commit the crime. If, in such a scenario, investigators find biological trace material (e.g. blood) on the victim's body that matches (e.g. the DNA profile of) the suspect, that finding constitutes evidence that supports the prosecution hypothesis more strongly than it does the defence hypothesis. If the investigators also find biological trace material on the suspect's body that matches the victim's, that finding constitutes another piece of evidence that supports the prosecution hypothesis more strongly than it does the defence hypothesis. However, the second piece of evidence is not independent to the first. A purely symbolic argumentation approach would struggle to specify how much more support the second piece of evidence adds to the first. But, a probabilistic approach, such as the one taken by Aitken et. al. can.

The DAG of the cross-transfer model is depicted in Figure 1. The root nodes G, B_v and B_s of the DAG are employed to specify hypotheses and facts. For example, $h_d = \overline{g}$ (the suspect is not guilty) is an example of a defence hypothesis with respect to the trace evidence in question and $h_p = g$ (the suspect is guilty) is an example of a prosecution hypothesis. If the victim's background (e.g. profession) puts him/her into contact with the type of trace material under analysis and the suspect's background does not, then b_v and $\overline{b_s}$ can be treated as undisputed observations. The leaf nodes of the DAG correspond to the available evidence. For example, retrieval of trace material from the victim matching the suspect and the failure to match trace material retrieved from the suspect with the victim could be represented as rm_v and $\overline{rm_s}$. In the evidential reasoning scenario considered herein, the value of the two pieces of evidence would be computed by the following two likelihood ratios:

$$LR(rm_v|h_p,h_d) = \frac{P(rm_v|g,b_v,\overline{b_s})}{P(rm_v|\overline{g},b_v,\overline{b_s})} \quad \text{and} \quad LR(\overline{rm_s}|h_p,h_d) = \frac{P(\overline{rm_s}|g,b_v,\overline{b_s})}{P(\overline{rm_s}|\overline{g},b_v,\overline{b_s})}$$

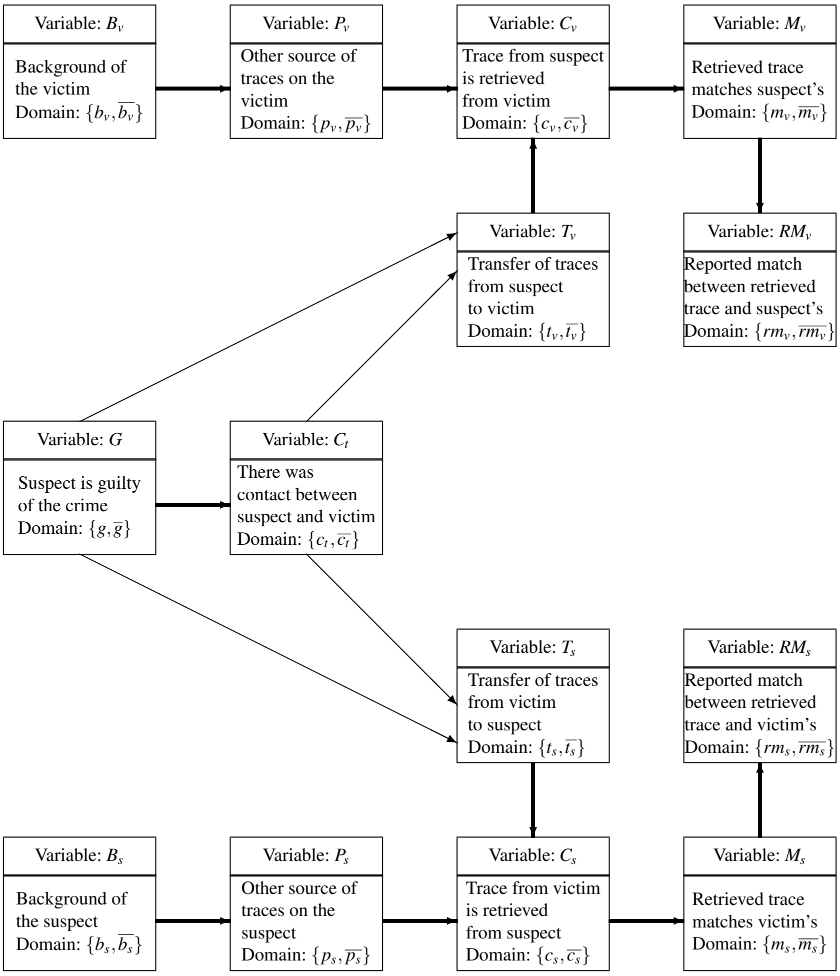


Fig. 1 Structure of a variation of Aitken, Taroni and Garbolino's cross-transfer Bayesian network. The arrow $T_v \rightarrow T_s$ of the original network has been omitted from this version such that the same conditional probability table can be used for both T_v and T_s and simplify the development of complete worked example based on this network.

The remainder of the DAG shown in Figure 1 consists of the other variables that affect the probability distributions of the hypothesis and evidence variables and conditional independence relationships between them, represented in the manner described above. For example, the variable C_t represents the proposition that there was contact between victim and suspect. Aitken et. al. chose to condition C_t on the hypothesis variable G. The variable T_v represents the proposition that there was a transfer of biological trace material from suspect to victim. T_v has both G and C_t as its parents. As such, G and T_v are not deemed to be probabilistically independent given C_t because, according to Aitken et. al., the probability distribution of T_v is still affected by the value of G even if C_t is known. Conversely, hypothesis variable G and

G	g		$\overline{g}$	
C_t	c_t	$\overline{c_t}$	c_t	$\overline{c_t}$
$P(t_s\|\text{parent values})$	p_1	p_2	p_3	p_4
$P(\overline{t_s}\|\text{parent values})$	$1-p_1$	$1-p_2$	$1-p_3$	$1-p_4$

Table 1 Conditional probability table of T_s

evidence variable RM_v are deemed conditionally independent given T_v, because, according to the author of the model, the probability distribution of retrieval of trace material matching the suspect from the victim is not affected by changes in the probability of the guilt hypothesis if it is already known whether biological trace material was transferred from suspect to victim. The entire network is constructed by incrementally considering the conditional independence relationships between variables in this way.

Once the DAG of the BN has been constructed, conditional probability distributions can be defined. As all variables in this BN are discrete, these probability distributions can be specified in the form of a conditional probability table, such as the one shown in Table 1 for variable T_s. This table contains a probability distribution for T_s for each combination of value assignments of the parent variables of T_s in the DAG: G and C_t. As such, there are 4 probability distributions for T_s, each defined by a pair of probabilities p_i, $1-p_i$.

3.1.3 Knowledge Acquisition Challenges

The development of accurate probabilistic models to apply the likelihood ratio approach to is challenging. The use of BNs aids the knowledge acquisition process somewhat. In particular, the conditional independence model represented by a BN's DAG facilities the inclusion of a larger number of variables, because the conditional independencies limit the number of variables that the probability distribution of each variable in the model needs to be conditioned on. While the specification of conditional independence relationships in the form of a BN's DAG is not trivial, the DAG has clear and generally understood meaning. Any given BN DAG can be validated, e.g. through peer evaluation, or even cross examination in court.

Various authors have developed template BNs to evaluate specific types evidential reasoning problem. Aitken et. al.'s cross-transfer model is one example of such a model. Other examples include BNs for assessing the value of two or more pieces of trace evidence (i.e. the two or multiple trace problem) [29], BNs for assessing the value of partial DNA matches in mixtures of DNA material [37] and BNs for evaluating the value of finding potential traces of certain accelerants in fire incidents [9]. Recently, various authors have shown how the validation of evidence, a concern that is typically tackled by means of argumentation based approaches [52], can be modelled by means of BNs [25,32]. The BN DAGs presented in such work, or parts thereof, can be reused and adapted to assess the value of evidence in practical problems.

Once the DAG of a BN is specified, conditional probability distributions for all variables must be defined. This is particularly challenging for those variables whose probability distributions are subjective, which means that the probability distributions are based on an expert's opinion or interpretation of data or domain theory. Experts find it particularly difficult to commit to precise numeric values as probabilities. Various approaches to allow experts to represent and reason with imprecise probabilities have been developed, such as qualitative probabilistic networks [20], semi-quantitative probabilistic networks [41] and credal networks [13]. The objective of such approaches is to enable experts to express both their knowledge and their lack of knowledge, thereby capturing subjective probabilities more accurately.

Even if they are expressed in an imprecise manner, it is usually difficult to assess the accuracy of subjective probability values by a person other than the expert who produced them, because subjective probabilities tend to refer to situations that occur infrequently. However, as illustrated by the discussion on R v T above, it is crucial that the accuracy of statements about subjective probability can be assessed by various people, such as the prosecution, the defence and the jurors. The remainder of this paper is concerned with a method to enable that.

3.2 A Method For Validating Subjective Probabilities

The previous subsection established the need for a method for validating subjective probabilities. As explained above, this approach must be able to capture the imprecision of an expert's knowledge. It must also be able to capture an expert's reasons for stating certain subjective probability values in a way that others can understand and attack, should they have reasons to do so. This subsection introduces such an approach.

3.2.1 Argumentation About Subjective Probability

In order to tackle the challenges set out above, this paper proposes an argumentation based approach to scrutinise subjective probability distributions. According to Schum, an argumentation based approach to evidential reasoning demonstrates that a piece of evidence supports a given hypothesis by (i) breaking down the inference steps between evidence and hypotheses into ones that can be tested and (ii) validating each inference step by subjecting it to a series of tests designed to discover flawed reasoning [52]. Schum illustrates this with an example of evaluating witness testimony of some observation that an event occurred. In this example, the argument pertaining to the credibility of a witness's testimony takes three steps, as follows:

Step 1:		the witness states that the event occurred
		the witness is truthful (veracity)
Step 2:	∴	the witness believes that the event occurred
		the witness is unbiased (objectivity)
Step 3:	∴	the witness senses give evidence that the event occurred
		the witness could observe what occurred (observational sensitivity)
	∴	the event occurred

The veracity, objectivity and observational sensitivity propositions in this sequence of arguments are then assessed by applying a series of test. In England and Wales, for example, the Turnbull guidelines [5] define a set of criteria (distance of observation, time of day, length of observation, etc.) that provide a basis for assessing an eyewitness's observational sensitivity. More recently, this approach has been formalised by means of argumentation schemes and critical questions [45,59]. Here, the argument schemes provide templates for valid inferences from one or more premises to a consequent/conclusion/claim, and the critical questions correspond to tests designed to discover ways in which the inferences might be defeated.

As Schum has argued, an argumentation driven evidential reasoning approach serves a different purpose than a Bayesian evidential reasoning approach [52]. Specifically, the former aims to assess to what extent a claim can be demonstrated from the evidence whereas the latter aims to assess how strongly evidence supports a hypothesis (independent of prior beliefs). Put differently, the former specifies to what extent reasoning holds up under scrutiny whereas the latter produces a numerical assessment of how relative likelihoods are affected by evidence. One can incorporate in a Bayesian model the same concerns introduced in argumentation models (such as veracity, objectivity and observational sensitivity) in order to render the Bayesian models more complete [32], or translate a Bayesian model to an argumentation model for validation [35]. But their purpose and output remain different.

Subjective probabilities, as used in Bayesian models, are inherently claims about the chances that situations occur under given conditions. Because such claims can affect the outcome of a likelihood ratio analysis, they must be scrutinised. Indeed, in England and Wales for example, the Crown Prosecutor must decide "whether evidence can be used in court and is reliable and credible" and "must be satisfied there is enough evidence to provide a realistic prospect of conviction against each defendant", before bringing a case to court [15]. As argued above, an argumentation approach is specified specifically for that purpose. But this raises the question how arguments can be formulated concerning claims concerning numerical values.

Earlier work has shown how significant features of conditional probability distributions can be expressed by means of constraints over the values in sets of probability distributions, such as those in a conditional probability table of a Bayesian network [20]. Constraints can represent qualitative and quantitative information with varying degrees of precision to be incorporated in a probabilistic model. For example, constraints can define how a probability $P(x|C)$ of proposition x, conditioned on a variable C with an ordered domain, changes with increasing or decreasing values of C and they can define upper and lower boundaries on probability values. The set of probabilities that satisfies a set of constraints believed to be true by an expert represents that expert's subjective probability. Therefore, constraints are employed as claims in argumentation models that provide the validatable support for the subjective probability the constraint entails. This implies that, in the type of models introduced in this work, an expert's argument based subjective probability is defined by the set of probabilities that satisfies the set of constraints entailed by the arguments that are believed to be true by the expert.

3.2.2 Constraint Problem and Analysis

In order to operationalise the intuitive definition of an expert's argument based subjective probability that was introduced in the previous subsection, it is necessary to introduce the concept of constraint satisfaction as it applies in this work. Specific types of constraint are introduced in Section 5 and an extended example is shown in 6. A conventional constraint problem [55] is specified by a tuple $\langle \mathbf{X}, \mathbf{D}, \mathbf{C} \rangle$, where

- $\mathbf{X}$ is a finite set of attributes $\{x_1, \ldots, x_n\}$,
- $\mathbf{D}$ is a finite set of domains $\{D_1, \ldots, D_n\}$ with $D_i = \{d_{i1}, \ldots, d_{in_i}\}$ for each attribute x_i, and
- $\mathbf{C}$ is a finite set of constraints, where each constraint $c \in \mathbf{C}$ over attributes $x_i, \ldots, x_j$ is a relation over the domains $D_i, \ldots, D_j$ (i.e. $c \subseteq D_i \times \ldots \times D_j$).

A set of assignments $\{x_1 : d_{1k_1}, \ldots, x_i : d_{ik_i}, \ldots, x_j : d_{jk_j}, \ldots, x_n : d_{nk_n}\}$ is said to satisfy a constraint c over the domains of $x_i, \ldots, x_j$ if $(d_{ik_i}, \ldots, d_{jk_j}) \in c$. A set of assignments is said to be a solution to the constraint problem $\langle \mathbf{X}, \mathbf{D}, \mathbf{C} \rangle$ if it satisfies all the constraints in $\mathbf{C}$.

The constraint problems derived from argumentation models that define conditional probability tables vary from conventional ones in two respects. Firstly, constraints are justified by arguments and only constraints justified by accepted arguments affect the solution space. Specifically, if $\mathbb{A}$ is the set of arguments in the argumentation model from which the constraints are derived, j is a function $\mathbf{C} \mapsto \mathbb{A}$ that maps each constraint to the argument that supports it, and $\mathbf{A} \subseteq \mathbb{A}$ is the set of arguments that are accepted by an individual, then a set of assignments is said to be a solution to the constraint problem if it satisfies all the constraints in the set

$$\mathbf{C}' = \{c \in \mathbf{C} | j(c) \in \mathbf{A}\}$$

The set $\mathbf{C}'$ is said to be the set of constraints admitted by $\mathbf{A}$. Secondly, the domains are not finite but the continuous interval $[0, 1]$. For the purposes of this application, it is necessary to find the entire solution space to the problem to allow local propagation of the probability values.

In general, the problem of finding the solutions to a conventional constraint problem is NP-complete [19]. Although the problems of interest tend to be small, in the sense that they involve a small number of variables, the number of constraint problem attributes is proportional to d^n where n is the number variables that the probability distribution is conditioned on and d is the number of values of each variable. In practice, the variables of probabilistic models often denote propositions or come with small domains. While larger, and even continuous, domains are possible, smaller domains tend to be more appropriate for variables that have subjective probability distributions because the vagueness/ambiguity of knowledge concerning these variables precludes larger, more precise, domains.

Solutions to these constraint problems can be computed in a number of ways. Conventional constraint problem solvers, such as MINION [28], can be employed provided an appropriate interface between the constraint solver and the argumentation system exists. This interface serves two roles. Firstly, a constraint problem must be defined for each set of accepted arguments, that is then solved from scratch. Secondly, the interface must discretise the domains. Such an approach can lead to very large solution spaces if the domains are discretised with fine granularity.

Dynamic constraint problems correspond to a sequence of conventional constraint problems, where each constraint problem is a minor variation of its predecessor [56].

Dynamic constraint problem solvers seek to improve on the performance of solving these conventional constraint problems in sequence with a conventional constraint solver by repairing the solution to one constraint problem with respect to the changes in specification to the next constraint problem. In other words, these techniques aim to reduce the required exploration of the search space for a new (set of) solution(s) after a change in the specification of a constraint problem, by ignoring those parts of the search space that could not possibly be affected by the change. Such techniques could be applied here to compute the implications of changes in the set of accepted arguments **A**.

Continuous constraint solvers, such as ECLiPSe [39] and RealPaver [31], are designed to handle constraint problems with continuous domains without requiring explicit discretisation of the domains in the problem specifications. They do so by computing approximations of the solution space rather than enumerating all individual solutions. For example, Sam-Haroud and Faltings have developed a method to find all solutions to a continuous constraint problem by means of a hierarchical decomposition of the solution space [51]. This approach has two important advantages with regard to solving the constraint problem defined above. Firstly, it allows for certain larger contiguous subspaces of the solution space that are either admitted or rejected by constraints in their entirety to be represented by means of a compact data structure. Secondly, it can represent a solution space with varying degrees of granularity. The former feature enables a large set of solutions to be represented and evaluated efficiently. The latter feature allows for the solver to scale its runtime and storage requirements (as well as its precision) to the available resource.

For the purposes of this paper, any of the aforementioned continuous constraint solvers can produce the spaces (or approximations thereof) of subjective conditional probabilities that satisfy the constraints admitted by a set of arguments **A** that are accepted by an expert. These must be used to compute intervals of probabilities $P_{\mathbf{A}}(X)$ of propositions X admitted by the probabilistic model under the constraints that are accepted by the set of arguments **A**. In general, finding $P_{\mathbf{A}}(X)$ is an ongoing problem being investigated in the area of interval probability theory [61]. However, probabilistic evidential reasoning problems tend to be modelled by means of Bayesian networks or simpler models that can be formalised by means of Bayesian networks. Credal networks provide a means to propagate convex sets probabilities adhering to the conditional independence relationships defined by means of a Bayesian network [13], thereby allowing the computation of $P_{\mathbf{A}}(X)$ once the space of conditional probability tables admitted by the constraints accepted by **A** have been found.

The ability to compute $P_{\mathbf{A}}(X)$ allows the system to answer three types of queries that pertain to the impact of the set of accepted arguments.

1. What is the value of evidence given the beliefs in **A**? As the value of evidence is computed by the likelihood ratio, the value of a piece of evidence e with regard to two hypotheses h_1 and h_2 given the beliefs in **A** is given by:

$$LR_{\mathbf{A}}(e|h_1,h_2) = \frac{P_{\mathbf{A}}(e|h_1)}{P_{\mathbf{A}}(e|h_2)} \tag{4}$$

   $LR_{\mathbf{A}}(e|h_1,h_2)$ is a convex set of likelihood ratios admitted by the argument set **A**.

2. What is the impact of rejecting an argument $\mathbf{a} \in \mathbf{A}$? This can be computed as the set of factors that convert the value of the evidence prior to rejecting a to the value of the evidence after rejecting a:

$$\frac{LR_{\mathbf{A}/a}(e|h_1,h_2)}{LR_{\mathbf{A}}(e|h_1,h_2)} \tag{5}$$

3. What is the impact of accepting a new argument $\mathbf{a} \notin \mathbf{A}$? Similar to query type 2, this is computed as the set of factors that convert the value of the evidence prior to accepting a to the value of the evidence after accepting a:

$$\frac{LR_{\mathbf{A}\cup\{a\}}(e|h_1,h_2)}{LR_{\mathbf{A}}(e|h_1,h_2)} \tag{6}$$

It is important to note that the valuations of evidence in (4), (5) and (6) result in intervals of likelihood ratios instead of single values, because they are computed on the basis of convex sets of probabilities. The sizes of the intervals express the degree of uncertainty that remains in assessing the value of evidence due to the imprecise nature of the subjective probabilities involved. All other things being equal, the more the subjective probabilities can be constraint by arguments, the smaller the intervals will tend to be. In court, any range of likelihood ratios that includes 1 should be interpreted to signify that the evidence does not provide clear support for one hypothesis over another. If the entire interval is below or above 1, both the the weakest level of support (i.e. the value closest to 1) and the size of the interval is important. The former specifies the weakest level of support that can be claimed and the latter the degree of doubt that remains in the mind of the expert.

3.2.3 Validating Subjective Probabilities

The previous subsection presented an approach to construct argument models that define subjective probability distributions of probabilistic evidential reasoning models and to compute the implications of arguments on the value of evidence. This subsection shows how such models facilitate validation of subjective probability distributions. To do so, it is necessary to examine how evidential reasoning models are constructed.

A forensic scientists who is to assess the value of a piece of evidence may produce a probabilistic model that enables him/her to do so. The methodology for producing such a model is well understood and involves: (i) defining two hypothesis propositions (defence and prosecution hypotheses, or null and alternative hypotheses) and the evidence proposition, (ii) identifying and defining the variables that affect the likelihood of the hypotheses, (iii) specifying the conditional independence relationships between the variables and (iv) specifying conditional probability distributions [12]. The argumentation based approach extends step (iv): where conditional probability distributions are subjective, the forensic scientist (iv.a) produces arguments supporting constraints over the conditional probability distributions, which are then (iv.b) solved to produce a set of solutions representing the scientists opinion.

The arguments are constructed by instantiating argumentation schemes. While argument schemes define valid forms of argument, their instantiations with respect to a particular case may not be. Each argument scheme is associated with a set of critical questions that is designed to uncover reasons why instantiations of that argument scheme may be invalid. Thus, the subjective probabilities used by the expert who basis his/her assessment of the value of evidence on a probabilistic model containing those subjective probabilities are validated in three stages. First, the expert must make the reasons for constraining subjective probabilities in certain ways explicit. As this involves instantiating argumentation schemes, the expert is required to identify how the prerequisite conditions for the claims made in this process are satisfied in his/her opinion. Next, the expert applies the critical questions of each argumentation scheme, with a view to assess whether the argument survives that test. Finally, an explicit model of an expert's reasoning in deciding subjective probabilities is available, in addition to the conventional probabilistic model and the assessment of the value of evidence it entails. This allows the expert's reasoning to be validated by others far more easily than a mere set of numbers in a probabilistic model would. In other words, it improves transparency. In a typical case, an expert's assessments concerning the value of a piece evidence, and by extension the associated argumentation models, may be scrutinised at various stages: by peers of the expert before releasing the outcome of an analysis, by the prosecutor prior to bringing the case to court and by the defence in court. Again, the validity of the conditions of arguments and the critical questions can be used as a guide in these subsequent validation processes.

3.3 Methodology For Defining Argumentation Schemes For Subjective Probabilities

The approach presented in this sections is crucially dependent upon the availability to a suitably broad range of argumentation schemes with associated critical questions. Although a range of argumentation schemes exist that are relevant to evidential reasoning [60], the use of argumentation schemes to construct arguments supporting constraints over conditional probability distributions has not been studied yet in much depth. The remainder of this paper is dedicated to defining such a set of argumentation schemes. This section discusses the methodology behind the definition of these argumentation schemes.

Some limitations must be imposed on the types of argumentation schemes that can be used. Because the purpose of this work is the validation of probabilistic models for assessing the value of evidence by experts, it is reasonable to assume that the arguments stem from individuals that aim to approach the assessment of evidence objectively and independently (even if they do not accomplish this). All claims concerning subjective probabilities stem from one or more sources, such as a data set or an expert's opinion. In a legal setting, such sources constitute evidence that affect the value assessment of the (direct) evidence. In other words, these sources are ancillary evidence. It is important that all such ancillary evidence used in an assessment of the value of evidence is catalogued and that its nature is identified because its reliability affects the reliability of the resulting assessment. Therefore, Section 4 introduces a set of argumentation schemes for so-called *source-based arguments* [60]. These source-based arguments interpret the ancillary evidence and justify claims concerning the nature of subjective probability distributions. These claims cannot necessarily be ex-

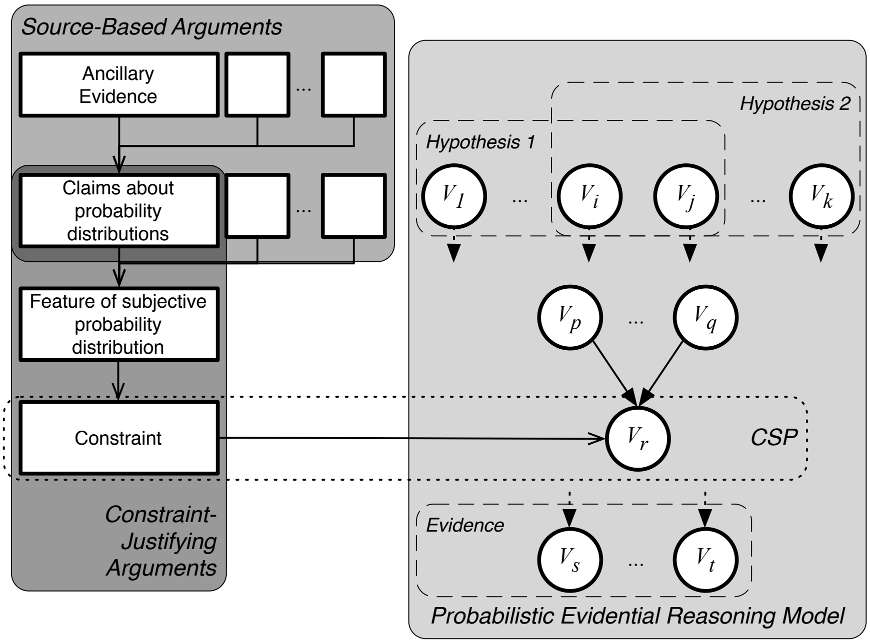


Fig. 2 Argumentation about Subjective Probabilities in Probabilistic Evidential Reasoning

pressed in the form of constraints. Therefore, a second set of *constraint-justifying argumentation schemes*, presented in Section 5, specify how such claims or combinations of such claims justify features of subjective probability distributions that can be expressed by means of constraints. This two-level argument model and its relation to a probabilistic model is presented in Figure 2.

Generally speaking, the source for a claim concerning a subjective probability distribution can be an expert's personal judgement, one or more data sets, a theory or set of first principles, or a combination of these sources [63]. Therefore, the source-based argumentation schemes of Section 4 represent the way claims concerning probability distributions can be derived from these sources and the critical questions cover the assumptions and potential causes for errors in interpreting the sources.

The constraint-justifying argumentation schemes are somewhat harder to identify as it is not feasible to capture all possible qualitatively significant features human experts may express about subjective probability distributions. The objective of this work is to define a sufficiently broad range of constraint-justifying argumentation schemes to represent and compute subjective probabilities in real-world probabilistic models for evidential reasoning. Over the past few decades, one strand of research in uncertainty in AI has devised a wide range of knowledge representation formalisms and associated inference mechanisms to represent and reason with subjective probabilities, by capturing very specific types of probabilities that correspond to a kind of features that is qualitatively significant and be identified by a human expert. An example of this is Adams' ε-semantics [1], which enables the definition of a logic to reason about propositions that have a probability close to 0 or 1. A separate strand of research in the same field aims to devise knowledge representation formalisms and inference mechanisms that are generic and can express any degree of uncertainty about probability. An example of this is Cozman's work on credal networks [13]. For the purposes of of this work, the former strand of research is reviewed in Section 5 to identify qualitatively significant features that may be exhibited by subjective probabilities and redefines them as argumentation schemes with associated sets of critical questions. In order to demonstrate that the range of argumentation schemes defined in this way is both usable and sufficiently broad, Section 6 uses them to define the subjective probabilities of a sophisticated probabilistic model from the forensic statistics literature by means of argumentation models.

3.4 Notation

The argumentation schemes defined in the following sections specify subjective conditional probability distributions of the form $P(C|\mathbf{B})$, where $\mathbf{B}$ is a set of variables. The set $\mathbf{B}$ is partitioned into a singleton containing B, a set $\mathbf{B_c}$ and a set $\mathbf{X}$ such that the probability distributions assume the form $P(C|B,\mathbf{B_c},\mathbf{X})$. The argumentation schemes presented below impose constraints over probabilities $P(C|B,\mathbf{b_c},\mathbf{X})$, where $\mathbf{b_c}$ is an assignment of values of the variables in $\mathbf{B_c}$. Specifically, the argumentation schemes entail constraints over the probability distribution of C under certain conditions specified by assignments of B, given that variables $\mathbf{B_c}$ are assigned the values $\mathbf{b_c}$ and for any assignment $\mathbf{x}$ of the variables $\mathbf{X}$ that meet the requirements of argumentation scheme. In what follows, the variable C is called the *consequent variable*, B is called the *antecedent variable* and $\mathbf{b_c}$ is called the *context* of the argumentation scheme.

4 Source-Based Argumentation Schemes & Critical Questions

This section presents a range of source-based argumentation schemes produced by means of the methodology of Section 3.3.

4.1 Argument from Expert Opinion

As explained above, expert opinion is an important source of ancillary evidence for subjective probability. As such, this section introduces a scheme for an argument from an expert's opinion about conditional probability distributions as ancillary evidence. Argumentation scheme 1 is a variation of Walton's argument from expert opinion [58] The key difference between Walton's argument from expert opinion and argumentation scheme 1 is that the generic notion of an assertion expressing expert opinion in the former is substituted by more specific propositions concerning the variables in $\{C,B\}\cup\mathbf{B_c}\cup\mathbf{X}$ that manifests itself as a feature F of the probability distributions $P(C|B,\mathbf{B_c},\mathbf{X})$.

Argumentation scheme 1: *Argument from Expert Opinion*

- E is an expert in domain D
- The variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X}$ and their interrelationships are a part of domain D
- E asserts that the conditional probability distributions $P(C|B,\mathbf{b_c},X)$ possess feature F

$\therefore$ The probability distribution $P(C|B,\mathbf{B_c},\mathbf{X})$ may be constrained by feature F

This more specific notion of an assertion allows for a more specific set of critical questions:

Critical Questions for argumentation scheme 1

1. *Expertise Question*: Is E credible as an expert in the relevant domain
   (a) Is E credible as an expert in D?
   (b) Is it credible for E to have an opinion concerning each of the variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X}$?
   (c) Is it credible for E to have an opinion on the interrelationship between the variable in $B \cup \mathbf{B_c}$ with C?
   (d) Is it credible for E to consider all relevant circumstances $\mathbf{x}$ that can be described as combinations of assignments of the variables in $\mathbf{X}$?
2. *Trustworthiness Question*: Is E trustworthy as an expert in D?
   (a) Is E biased?
   (b) Is E honest?
   (c) Is E conscientious?
3. *Backup Evidence Question*: Is E's assertion based on evidence?
4. *Domain Question*: Is F in the relevant domain?
   (a) Is F in domain D?
   (b) Is it reasonable to classify the variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X}$ as belonging to domain D?
5. *Consistency Question*: Is the assertion that the conditional probability distribution possess feature F consistent with assertions of other experts in D?
6. *Opinion Question*: Did E's assertion imply feature F?

This set of six critical questions cover the same issues as the critical questions associated with Walton's argument from expert opinion: expertise, trustworthiness, backup evidence, domain, consistency and opinion. Here though, the expertise and domain questions are elaborated because of the use of a more specific notion of the assertion.

4.2 Argument from Data Set

Quantitative information about probability distributions typically stems from data sets. Increasingly, forensic databases can provide information about the frequency of occurrence of certain characteristics of objects potentially related to case: e.g. paint traces, glass fragments and shoe prints [10]. Argumentation scheme 2 aims to specify the structure of an argument from a data set. It is a source-based argumentation scheme that is similar to argumentation scheme 1. However, instead of qualifying the domain D that the data set S provides information about, the scheme specifies the variables about which the data set provides information.

Argumentation scheme 2: *Argument from Data Set or Model*

- S is a data set or model that relates the variables in $\{B\} \cup \mathbf{B_c} \cup \mathbf{X}$ with C
- S implies a feature of F concerning the probabilities $P(C|B, \mathbf{b_c}, X)$

$\therefore$ The probability distributions $P(C|B, \mathbf{B_c}, \mathbf{X})$ may be constrained by feature F

Deriving information related to probability distributions from a data set is a non-trivial task that can be prone to errors. Below is a list of critical questions that seek to identify common errors in the use of arguments from data sets. The scope questions assume the same role for argumentation scheme 2 that the expertise and domain questions assume with regard to argumentation scheme 1. Specifically, they aim to assess whether the data set provides information on the relevant variables. The representativeness question aims to determine whether the population of the data set accurately reflects the relevant features of the population under consideration in the evaluation of evidence. If the population under consideration is a subset of the population represented by the data set, then it may possess unique characteristics that affect the probability distribution. The precision question seeks to test whether the precision feature F is not unwarranted given the data set it is extracted from. The opinion and consistency questions assume the same role as in an argument from expert opinion.

Critical Questions for argumentation scheme 2

1. *Scope Question*: Does data set S provide the necessary information to inform the probability distributions $P(C|B, \mathbf{B_c}, \mathbf{X})$?
   (a) Does data set S cover the variables in $\{C, B\} \cup \mathbf{B_c}$?
   (b) Does data set S cover the interrelationships between the variables in $B \cup \mathbf{B_c}$ with C?
   (c) Does data set S cover all variables in $\mathbf{X}$ necessary to identify the relevant circumstances covered by feature F?
   (d) Does data set S cover all relevant circumstances $\mathbf{x}$ that can be described as combinations of assignments of the variables in $\mathbf{X}$?
2. *Representativeness Question*: Is the population considered in data set S representative for the population under investigation in the present case?
3. *Precision Question*: Is the volume and precision of data set S consistent with the precision of feature F?
4. *Opinion Question*: Does data set S entail feature F?
5. *Consistency Question*: Is the observation of feature F in data set S consistent with other data sets?

4.3 Argument from First Principles

Domain theories can entail a broad range of features of probability distributions. First principles can provide causal relationships between variables and qualitative diagnostic information. Where theories stem from or have been validated with empirical data, they can also provide quantitative information on probability distributions. Argumentation scheme 3 is a template for an argument from first principles. It employs the same structure as that of an argument from expert opinion, but the expert E has been replaced by a theory T.

Argumentation scheme 3: *Argument from First Principles*

- T is a widely accepted theory concerning domain D
- The variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X}$ and their interrelationships are a part of domain D
- T implies feature F concerning the probabilities $P(C|B,\mathbf{b_c},X)$

$\therefore$ The probability distribution $P(C|B,\mathbf{B_c},\mathbf{X})$ may be constrained by feature F

The validation of a theory as a source of a feature of probability distributions is somewhat different to the validation of expert opinion. The representativeness questions test whether the theory can be applied to the circumstances of the case and to answer the questions at hand. The scope questions test whether the theory covers all the variables of interest. The trustworthiness question aims to assess whether the experts who formulated T can be trusted. Finally, the consistency questions determine whether the theory is consistent with views held in domain D.

Critical Questions for argumentation scheme 3

1. *Representativeness Question*: Is T applicable to the case?
   (a) Is the population under consideration for computing the conditional probability distribution representative of the population to which the theory is applicable?
   (b) Is the theory applicable to the scenarios under consideration?
   (c) Does theory T cover features such as F?
2. *Scope Question*: Does T provide the necessary information to inform the probability distributions $P(C|B,\mathbf{B_c},\mathbf{X})$?
   (a) Does theory T cover the variables in $\{C,B\} \cup \mathbf{B_c}$?
   (b) Does theory T cover the interrelationships between the variables in $B \cup \mathbf{B_c}$ with C?
   (c) Does theory T cover the relevant circumstances of the case, as they can be represented by the variables in $\mathbf{X}$?
3. *Trustworthiness Question*: Are the people who propose T trustworthy?
4. *Backup Evidence Question*: Is T assertion based on evidence?
5. *Consistency Question*: Is T consistent with the state of knowledge in the domain?
   (a) Is T a generally accepted theory by experts in D?
   (b) Are there theories in D that conflict with T?
6. *Opinion Question*: Does T entail feature F?

4.4 Argument from Hybrid Sources

Conditional probability distributions do not necessarily stem from one type of source only. Arguments from data or first principles often involve expert judgement in the selection of the relevant data set or first principles and in their interpretation. This subsection identifies how argumentation schemes can be composed to produce such arguments from hybrid sources.

4.4.1 Expert Judgement in Source Selection

An argument from a data set or first principles often relies on expert judgement in the selection of the data set or first principles to be applied. The arguments made in such a case can be represented by combining the above argumentation schemes for arguments from a data set and arguments from first principles with Walton's argument from expert opinion. An example can illustrate this. Consider a situation where an expert proposes that a data set S is suitable to inform probability distributions of C conditioned on $\{B\} \cup \mathbf{B_c} \cup \mathbf{X}$. This can be formulated as an argument from expert opinion (in Walton's generic form):

- E is an expert in domain D
- E asserts that S is a data set that relates the variables in $\{B\} \cup \mathbf{B_c} \cup \mathbf{X}$ with C
- The selection of a suitable data set concerning the variables in $\{C, B\} \cup \mathbf{B_c} \cup \mathbf{X}$ is within E's area of expertise D

$\therefore$ S is a data set that relates the variables in $\{B\} \cup \mathbf{B_c} \cup \mathbf{X}$ with C (*)

The consequent of this argument has the form of the antecedent of an argument from a data set (where the link is identified by means of a (*) symbol):

- S is a data set that relates the variables in $\{B\} \cup \mathbf{B_c} \cup \mathbf{X}$ with C (*)
- S implies a feature of F concerning the probabilities $P(C|B, \mathbf{b_c}, X)$

$\therefore$ The probability distributions $P(C|B, \mathbf{B_c}, \mathbf{X})$ may be constrained by feature F

4.4.2 Argument from Expert Judgement

Even if data sets or first principles fail to meet the requirements of sections 4.2 and 4.3 respectively, they may still inform expert opinion. Based on such data sets and/or first principles, an expert may produce a conservative estimate of the relationships that exist between variables and formulate a constraint accordingly. To represent such arguments, arguments from data sets, arguments from first principles or a combination of both can be constructed that specify features on the basis of which an expert formulates a conservative estimate. This requires a new argumentation scheme for an argument from expert judgement:

Argumentation scheme 4: *Argument from Expert Judgement*

- E is an expert in domain D
- The variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X}$ and their interrelationships are a part of domain D
- The probability distributions P_1 is constrained by feature F_1

$\vdots$

- The probability distributions P_n is constrained by feature F_n
- E asserts that *if* the probability distribution P_1 is constrained by feature F_1 *and* ... *and* the probability distribution P_n is constrained by feature F_n, *then* the probability distributions $P(C|B,\mathbf{B_c},\mathbf{X})$ possess feature F

$\therefore$ The probability distribution $P(C|B,\mathbf{B_c},\mathbf{X})$ may be constrained by feature F

The application of this scheme is best illustrated by means of an example. Consider a situation where a data set S relates the variables in $\{B\} \cup \mathbf{Y}$ with C, but where conditional probability distributions $P(C|B,\mathbf{Y},\mathbf{X})$ need to be specified. In this case, the following argument from a data set can be formulated:

- S is a data set that relates the variables in $\{B\} \cup \mathbf{Y}$ with C
- S implies a feature of F concerning the probabilities $P(C|B,\mathbf{b_c},Y)$

$\therefore$ The probability distributions $P(C|B,\mathbf{Y})$ may be constrained by feature F (A)

A feature F that affects probability distributions of the form $P(C|B,\mathbf{Y})$ does not necessarily generalise to probability distributions of the form $P(C|B,\mathbf{Y},\mathbf{X})$. The reason for this is that a constraint that is valid in any set of circumstances defined by an assignment $\mathbf{y}$ of the variables in $\mathbf{Y}$ is not necessarily valid in each of the more specific circumstances that can be defined by an assignment $\mathbf{x}$ of the variables in $\mathbf{X} \cup \mathbf{Y}$. However, an expert may argue, for example, that the variables in $\mathbf{X}$ do not affect the probability of C. This judgement entails the feature F expressed in proposition (A). Therefore, the following argument would apply in this example:

- E is an expert in domain D
- The variables in $\{C,B\} \cup \mathbf{B_c} \cup \mathbf{X} \cup \mathbf{Y}$ and their interrelationships are a part of domain D
- E asserts that a feature of type F concerning the probability distributions describing the relationship between variables B and C given circumstances described by $\mathbf{Y}$ generalise to all possible circumstances that can be described by the variables in $\mathbf{X} \cup \mathbf{Y}$ (B)

$\therefore$ The probability distributions $P(C|B,\mathbf{Y},\mathbf{X})$ may be constrained by feature F

Here, note that proposition (B) can be reformulated as "E asserts that *if* the probability distribution $P(C|B,\mathbf{Y})$ is constrained by feature F *then* conditional probability distributions $P(C|B,\mathbf{Y},\mathbf{X})$ possess feature F", thereby assuming the form of the final antecedent of argumentation scheme 4.

5 Constraint-Justifying Argumentation Schemes & Critical Questions

The previous section introduced a number of argumentation schemes that justify claims about of probability distributions. In this section, schemes are introduced that identify ways of translating these claims into specific features that correspond to constraints. The constraints can operationalise the features at different levels of precision. As such, this section is organised into three subsections corresponding to different categories of precision.

In the process of justifying constraints, a number of decisions are made. These decisions take the form of additional propositions in the antecedents of the argumentation schemes. The critical questions associated with argumentation schemes introduced in this section largely aim to validate these new decisions. Where this is the case, the critical questions are self-explanatory and no further explanations are provided. As is shown, argumentation schemes representing constraints of higher precision tend to involve a larger set of decisions and therefore a broader range of critical questions.

5.1 Arguments for Qualitative Features

This subsection introduces a number of argumentation schemes that justify common but weak relationships over probability values. These relationships are weak because they imply constraints that are not particularly restrictive. However, this makes them relatively uncontroversial and easy to commit to by experts.

5.1.1 Arguments for Qualitative Influences

Qualitative influences stem from qualitative certainty networks [43], an abstraction of a variety of probabilistic models. In their most basic form, qualitative influences constrain probability distributions of the form $P(C|B,\ldots)$, where B and C are variables with totally ordered domains. Intuitively speaking, a variable B is said to positively (negatively) influence C if an increase in the probability of higher values of B implies an increase (decrease) in the probability of higher values of C. argumentation scheme 5 specifies the conditions for a positive (negative) qualitative influence for variables with Boolean domains (i.e. propositions).

Argumentation scheme 5: [*Positive*|*Negative*] *Qualitative Influence*

- B has a [non-negative|non-positive] effect on the likelihood of C

$\therefore$ $P(c|b,\mathbf{x})[\geq | \leq]P(c|\overline{b},\mathbf{x})$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 5

1. Is the direction of the influence correct?
2. Is the direction of influence dependent on context assumptions?
   (a) Are there assignments $\mathbf{x}$ of the variables in $\mathbf{X}$ where the direction of the effect of B on the likelihood of C is reversed?

(b) Are there circumstances where the direction of the effect on B on the likelihood C is reversed?

In the cross-transfer model, it can be argued that G has a non-negative effect on C_t. A justification for this argument is that the type of crime scenario described by g is enabled by contact between suspect and victim c_t. More precisely, certain variants of scenarios involving g require contact (c_t). It can also be argued that P_v (P_s) has a non-positive effect on C_v (C_s). A justification for this argument is that the presence of another source of the same type of trace material being investigated as evidence (p_v or p_s) reduces the proportion of trace material related to the crime (if any is present) and, therefore, the probability of recovering it (c_v or c_s).

Renooij et. al. [49] have extended this concept of qualitative influences that are valid in all circumstances with one that is defined to be applicable to certain pre-specified circumstances only. Context-specific qualitative influences come with a context defined by a set of value assignments of variables the probability distributions are conditioned on. They constrain the probability distributions that satisfy the context only.

More formally, argumentation scheme 6 defines a context-specific qualitative influence of a variable B on a variable C, where the probability distribution of C is conditioned on B and a set of variables $\mathbf{B_c}$. The qualitative influence defined in the argumentation scheme is only deemed valid in those situations where the variables in $\mathbf{B_c}$ take the set of variable-value assignments $\mathbf{b_c}$. $\mathbf{b_c}$ is said to be the context of the qualitative influence.

Argumentation scheme 6: [*Positive*|*Zero*|*Negative*] *Qualitative Influence*

- B has a [non-negative|no|non-positive] effect on the likelihood of C given an assignment b_c of the variables in B_c

---

$\therefore$ $P(c|b,\mathbf{b_c x})[\geq \mid = \mid \leq]P(c|\overline{b},\mathbf{b_c x})$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 6

1. Is the direction of the influence correct?
2. Is the direction of influence dependent on context assumptions not included in $\mathbf{b_c}$?
   (a) Are there assignments $\mathbf{x}$ of the variables in $\mathbf{X}$ where, given the context $\mathbf{b_c}$, the direction of the effect of B on the likelihood of C is reversed?
   (b) Are there circumstances where, given the context $\mathbf{b_c}$, the direction of the effect on B on the likelihood C is reversed?

Note that scheme 6 is scrutinised by the same set of critical questions as those of scheme 5, but that the scrutiny is specific to the context of interest.

In the cross-transfer model, it can be argued that the possible existence of another source of trace material on the victim (P_v) has no effect on the probability of retrieving trace material from the victim's body belonging to the suspect (C_v) if there has been no transfer of trace material from suspect to victim ($\overline{t_v}$). This can be justified by an argument that:

i. P_v can dilute the amount of suspect related trace material transferred to the victim only, and
ii. t_v is the only source of suspect related trace material transferred to the victim.

5.1.2 Arguments for Combined Effects: Additive Synergy

Qualitative influences describe the impact of the probability distribution of one variable on another. Synergies specify how different influences with a common consequence relate to one another. In other words, they describe qualitatively significant features of "intercausal reasoning" [62]. Synergies involve three variables B_1, B_2 and C, such that there exist qualitative influences from B_1 to C and from B_2 to C. In the literature on qualitative probabilistic reasoning, two types of synergies have been identified: additive synergies and product synergies.

As defined in argumentation scheme 7, additive synergies express how the combined effect of two causes compares to the sum of the constituent effects. If the combined effect of the two antecedent variables is greater than the sum of the individual effects, then there is said to be a positive additive synergy between the influences of B_1 and B_2 on C. Conversely, if there is a negative additive synergy, then the combined effect of the parent variables is smaller than the sum of the constituent effects.

Argumentation scheme 7: [*Positive*|*Negative*] *Additive Synergy*

- b_1 has a non-negative effect on the likelihood of c
- b_2 has a non-negative effect on the likelihood of c
- the combined effect of b_1 and b_2 on the likelihood of c is [greater|smaller] than the sum of the constituent effects of b_1 and b_2

$\therefore$ $P(c|b_1,b_2,\mathbf{x}) + P(c|\overline{b_1},\overline{b_2},\mathbf{x})$ $[\geq | \leq]$ $P(c|\overline{b_1},b_2,\mathbf{x}) + P(c|b_1,\overline{b_2},\mathbf{x})$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 7

1. Does b_1 have a non-negative (non-positive) effect on the likelihood of c (tested by the critical questions of Argumentation scheme 5)?
2. Does b_2 have a non-negative (non-positive) effect on the likelihood of c (tested by the critical questions of Argumentation scheme 5)?
3. Is the underlying feature an additive synergy?
   (a) Does the underlying feature concern an effect on the likelihood of c?
   (b) Does the underlying feature compare the effect of the influences isolation of one another with the effect of the combined influences?
4. Is the direction of the additive synergy correct?
   (a) Does a change in the likelihood of b_1 (b_2) have a [greater|lesser] effect on the likelihood of c with higher values of b_2 (b_1)?
   (b) Are there assignments $\mathbf{x}$ of the variables in $\mathbf{X}$ where the direction of the additive synergy is not [positive|negative]?

Note that argumentation scheme 7 specifies non-context-specific positive qualitative influences of B_i on C. If either b_1 or b_2 has a non-positive effect on the likelihood of c, but not both, then the inequation signs need to be reversed (from $[\geq \mid \leq]$ to $[\leq \mid \geq]$). If both b_1 and b_2 have a non-positive effect on the likelihood of c, then the constraint applies as it is introduced in scheme 7. The scheme can also be generalised to context-specific influences by restricting the constraints to the relevant contexts.

In the cross-transfer model, it can be argued that the guilt hypothesis g and contact between suspect and victim c_t both have a non-negative effect on the transfer of traces from suspect to victim t_v. In the scenarios that Aitken et. al. have in mind, contact between suspect and victim in the context of the guilt scenario is that of a violent struggle. Specifically, the authors have cases of alleged rape or homicide in mind [3]. It is reasonable to argue that the transfer of trace material between two people is more likely in a violent struggle as part of a violent crime, than it is in most (albeit not all) scenarios of contact between this two people. This implies that the combined effect of g and c_t on t_v is greater than the sum of the constituent effects and, hence, that there is a positive additive synergy between the two effects.

5.1.3 Arguments for Explaining In/Away: Product Synergy

Intuitively, a product synergy describes whether a pair of explanations concerning the same consequent variable reinforce one another ("explaining in") or undermine one another ("explaining away"). As defined in argumentation scheme 8, given a qualitative influence of B_1 on C and of B_2 on C, a positive product synergy describes how an increase in the likelihood of b_1 (b_2) affects the likelihood of b_2 (b_1). As with additive synergies, the argumentation scheme can be extended to accommodate negative qualitative influences and context-specific ones.

Argumentation scheme 8: [*Positive*|*Negative*] *Product Synergy*

- b_1 has a non-negative effect on the likelihood of c
- b_2 has a non-negative effect on the likelihood of c
- the explanations b_1 and b_2 [reinforce|undermine] one another

$\therefore$ $P(c|b_1,b_2,\mathbf{x})P(c|\overline{b_1},\overline{b_2},\mathbf{x})\ [\geq \mid \leq]\ P(c|\overline{b_1},b_2,\mathbf{x})P(c|b_1,\overline{b_2},\mathbf{x})$

Critical Questions for argumentation scheme 8

1. Does b_1 have a non-negative (non-positive) effect on the likelihood of c (tested by the critical questions of Argumentation scheme 5)?
2. Does b_2 have a non-negative (non-positive) effect on the likelihood of c (tested by the critical questions of Argumentation scheme 5)?
3. Is the underlying feature a product synergy?
   (a) Does the underlying feature concern the plausibility of b_1 and b_2 as explanations of c?
   (b) Does the underlying feature compare the plausibility of b_1 (b_2) in the absence of b_2 (b_1) with the plausibility of b_1 (b_2) in the presence of b_2 (b_1)?
4. Is the direction of the product synergy correct?

(a) Does an increase in the likelihood of b_1 (b_2) [increase|decrease] the likelihood of b_2 (b_1)?
(b) Does a decrease in the likelihood of b_1 (b_2) [decrease|increase] the likelihood of b_2 (b_1)?
(c) Are there assignments **x** of the variables in **X** where the direction of the product synergy is not [positive|negative]?

In the cross-transfer model, an additional edge $T_v \rightarrow T_s$ or $T_s \rightarrow T_v$ may be added to model correlation between transfers of trace material in either direction for known values of C_t. Consider an alteration of the model where there are three edges to T_s: $G \rightarrow T_s$, $C_t \rightarrow T_s$ and $T_v \rightarrow T_s$[1]. In this version, the context that there has been contact between victim and suspect, allows two remaining explanations for trace transfer from victim to suspect t_s:

g : the scenario where the suspect is guilty of the crime under consideration, and
t_v : transfer of traces from suspect to victim.

One could argue that these explanations ought to reinforce one another: i.e. g is more likely an explanation for t_s given t_v and t_v is more likely an explanation for t_s given g.

5.2 Arguments for Extreme Probabilities

Probabilistic models designed for evidential reasoning with legal and forensic applications commonly feature probabilities that are equal or fairly close to 0 or 1. Such extreme probabilities are important to probabilistic reasoning in general, so much so that a number of authors have devised specific qualitative calculi to represent and reason with very large and very small probability values. Examples include Adams' ε-semantics [1,44], Goldszmidt and Pearl's system-Z+ [30] and Spohn's ordinal conditional functions [53]. The reasons that justify incorporating extreme probabilities in conditional probability tables vary considerably. This section identifies some of the most important ones.

5.2.1 Arguments for Sufficient and Necessary Conditions

In a probabilistic model, sufficient and necessary are relevant to any model containing a variable C conditioned on one or more variables **B** such that an assignment **b** of the variables in **B** constitutes a sufficient or necessary condition for C to assume a particular value. In this section, argument schemes for sufficient and necessary conditions are defined for variables C that represent a proposition (though the idea is easily extended to variables with larger domains). If **b** is a sufficient condition, it is known that c is always true if **b** is true, and hence that $P(c|\mathbf{b}) = 1$. This is specified in argumentation scheme 9. In the cross-transfer example, it can be argued that the transfer to trace material from suspect to victim (t_v) and the absence of a secondary source of trace material of the same type (p_v) is a sufficient condition for a forensic examiner

[1] This version of the model is not included in Figure 1 to make the model easier to understand as the basis for demonstrating this work.

to retrieve trace material from the victim belonging to the suspect (c_v) because "there is no biological material from elsewhere" [3].

Argumentation scheme 9: *Arguments for Sufficient Conditions*

- **b** is a sufficient condition for c

$\therefore$ $P(c|\mathbf{b},\mathbf{x}) = 1$

Critical Questions for argumentation scheme 9

1. Might **b** be a necessary condition rather than a sufficient condition?
2. Is it possible for **b** to be true and c to be false?

If **b** is a necessary condition, it is known that c is always false if **b** is false, and hence that $P(c|\neg\mathbf{b}) = 0$. This is specified in argumentation scheme 10. The cross-transfer model contains a number of examples of necessary conditions. Contact between suspect and victim (c_t) is a necessary condition for the transfer of trace material both from suspect to victim (t_v) and from victim to suspect (t_s), because according to Aitken et. al. "if there is no contact, it is assumed that there is no transfer (i.e. no secondary transfer)" [3]. These transfers of trace material (t_v and t_s) are themselves necessary conditions for the retrieval of transferred traces from victim or suspect (c_v and c_s) because no crime-related trace material belonging to the suspect/victim could be retrieved from the victim/suspect if there was no transfer of trace material between both [3].

Argumentation scheme 10: *Arguments for Necessary Conditions*

- b is a necessary condition for c

$\therefore$ $P(c|\bar{b}) = 0$

Critical Questions for argumentation scheme 10

1. Might b be a sufficient condition rather than a necessary condition?
2. Is it possible for b to be false and c to be true?

5.2.2 Arguments for Definitions

If **b** is both a sufficient and necessary condition for c, it is said that **b** defines c. This situation is described in argumentation scheme 11.

Argumentation scheme 11: *Argument from Definition*

- c if and only if **b**

$\therefore$ $P(c|\mathbf{b}) = 1$
$P(c|\neg\mathbf{b}) = 0$

Critical Questions for argumentation scheme 11

1. Is each $b \in \mathbf{b}$ a necessary condition for c?
2. Is **b** a sufficient condition for c?
3. Is it possible for **b** to be true and c to be false?
4. Is it possible for a $b \in \mathbf{b}$ to be false and c to be true?

In the cross-transfer model, the presence of trace material from another source of the same type as that under investigation (p_v and p_s) on the victim/suspect is defined in terms of the background of the victim and the suspect (B_v and B_s). The variables B_v and B_s provide a means to explain the source of the presence of non-crime related trace-material and formulate defence and prosecution hypotheses.

5.2.3 Arguments for Observations/Classifications

A common feature of probabilistic models of evidential reasoning is that they may contain variables representing the observation or classification of a circumstance or set of circumstances denoted by other variables. As with an ”Argument from Definition”, an ”Argument from Observation/Classification” requires a particular situation defined by a tuple of value-assignments **b** of the variables in **B** that ought to be observed or classified as c. Unlike the application of a sufficiently precise definion, however, the acts of observation and classification are subject to errors. Typically, there is a small probability α that a situation that does not satisfy **b** is incorrectly observed or classified as c (aka Type I error) and a small probability β that a situation that does satisfy **b** is incorrectly observed or classified as not satisfying c (aka Type II error). Argumentation scheme 12 defines such an Argument from Observation/Classification.

Argumentation scheme 12: *Argument from* $[Observation|Classification]$

- The condition defined by **b** is [observed|classified] as c
- The type I error of classifying c is less than or equal to α

$\therefore$ $P(c|\neg\mathbf{b}) \leq \alpha$

- The condition defined by **b** is [observed|classified] as c
- The type II error of classifying c is less than or equal to β

$\therefore$ $P(c|\mathbf{b}) \geq 1-\beta$

Critical Questions for argumentation scheme 12

1. Is the relationship between the variables in **B** and C one of [observation|classification]?
   (a) Can c be fully explained by **b** or type I error? Are there situations other than **b** or type I error that explain c?
   (b) Can $\overline{c}$ be fully explained by the absence of **b** or type II error? Are there situations other than the absence of **b** or type II error that explain $\overline{c}$?
2. Is there evidence to support the type I error? Is there evidence to support the type II error?
3. Are there reasons to believe that the probability of type I error is greater than α? Are there reasons to believe that the probability of type II error is greater than β?

In the cross-transfer model, this argumentation scheme can be applied to the act of reporting trace matches, for trace evidence collected from the victim and the suspect.

More precisely, the reported trace matches on victim (rm_v) and suspect (rm_s) are observations of actual trace matches on victim (m_v) and suspect (m_s) respectively. In the process of observing the existence or absence of a match, Argumentation Scheme 12 enables the specification of an upper bound α on the probability that a match is reported where there is none (false positive) and an upper bound β on the probability that the absence of a match is reported where there is one (false negative).

5.2.4 Arguments for closed-world Assumptions

Intuitively, a closed-world assumption with regard to a proposition implies that all explanations for that proposition have been incorporated in our knowledge base or model. Thus, if none of the explanations in the model are applicable, the proposition is false. In other words, a closed-world assumption is an explicit statement that negation-as-failure can be applied in the model with respect to a given proposition [48].

In probabilistic models for evidential reasoning, the equivalent of committing to a closed-world assumption is to set the probability of a proposition to close to 0 in conditions where the model has no explicit reason to justify the proposition. Again, this is a common feature of a probabilistic model, but the corresponding assumption is often not accounted for. Argumentation scheme 13 provides a means for constructing arguments for making closed-world assumptions. Note that the argumentation scheme employs qualitative influences (context-specific or context-free ones) for describing reasons for accepting a proposition because qualitative influences are the weakest (least constraining) form of probabilistic justification for a proposition in this set of constraint-justifying argumentation schemes. It can be shown that all other arguments for increasing the probability of a proposition entail a qualitative influence.

Argumentation scheme 13: *Argument from closed-world Assumption*

- All reasonably plausible qualitative influences explaining c have been accounted for
- The probability of c occurring in the absence of any of the reasons accounted for less than or equal to ε

$\therefore$ $P(c|\mathbf{b}) \leq \varepsilon$ for any assignment $\mathbf{b}$ that does not constitute an explanation of c

Critical Questions for argumentation scheme 13

1. Is it possible to justify the absence of reasons for c other than those accounted with an error no greater than ε?
2. Is there any evidence to support the assertion that there are no alternative explanations for c other than those accounted for in the model?
3. Are there any explanations for c for which there is no corresponding qualitative influence or sufficient condition?
4. Is there any evidence to support that the error margin is not greater than ε?
5. In the absence of any explanation for c, is there any reason to believe that the probability of c could be greater than ε?

In the cross-transfer model, one might argue that a closed-world assumption could be made with regards to contact c_t and that the guilt scenario g is the most plausible explanation for c_t. This is not to say that contact is impossible in scenarios where the suspect is not guilty. In defining the closed-world assumption, a (conservative) upper bound must be selected representing the probability ε that there is a random contact between the suspect and the victim. A closed-world assumption is justified if it is reasonable to say that ε is low (e.g. Aitken et. al. use $\varepsilon = 0.01$ in this model). Whether this is a reasonable assumption depends on the circumstances of the victim and the suspect, and the critical questions seek to test this. For example, if victim and suspect were to live or be employed in relative proximity of one another, that would be a reasonable alternative explanation (tested by critical question 2), invalidating the closed-world assumption.

5.3 Arguments for Quantitative Features

This subsection presents a number of argumentation schemes that can express constraints that contain absolute quantities affecting the probability distributions. With such constraints, features of probability distributions can be expressed in greater precision. However, being more precise, such constraints tend to involve a greater number of decisions by the expert committing to the underlying argument. Therefore, the arguments tend to need to be subjected to larger sets of critical questions. Such arguments also tend to be more controversial.

5.3.1 Arguments for Boundary Conditions

Based on a data set, judgement or a combination thereof, an expert may argue to impose a boundary condition on the probability of a particular proposition under certain conditions. By means of argumentation scheme 14 an upper or lower bound of the conditional probability of a proposition c given $\mathbf{b}$, where $\mathbf{b}$ is an assignment of values to a subset of the variables that C is conditioned on, is defined and justified. The key implication of defining a boundary condition is that it holds for any combination of variable assignments of other variables that C is conditioned on, because the boundary condition is specified with regards to the conditional probability distribution in the probabilistic model.

Argumentation scheme 14: *Argument for Boundary Condition*

- The [lower|upper] bound on the probability of c given $\mathbf{b}$ is p

$\therefore$ $p[\leq \mid \geq]P(c|b,x)$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 14

1. Is it possible to infer that [a lower bound is greater than 0|an upper bound is smaller than 1] given the data at hand?
2. Is there any evidence to support the boundary condition?
3. Where the boundary condition relies on the relative occurrence of two events:

   (a) Are the two events related in such a way that a relative comparison is possible?
   (b) If not, are the absolute frequencies of the two events known?
4. Does a boundary condition hold for each of the possible circumstances that can be described by assignments $\mathbf{x}$ of $\mathbf{X}$?
5. Are there reasons to believe that the [lower|upper] bound is [smaller|greater] than p under certain circumstances? Can those circumstances be described by assignments $\mathbf{x}$ of $\mathbf{X}$?

5.3.2 Arguments for Order-of-Magnitude Influences

Qualitative influences as specified above define the direction of change of the probability of a proposition in response to a change in the probability of a second proposition that the first is conditioned on. Parsons has demonstrated how that concept can be extended to define magnitudes of change rather than just the direction of change [41]. In particular, he demonstrates the use of a relative order-of-magnitude calculus that expresses magnitudes by means of pairwise comparison constraints [17,46], and an absolute order-of-magnitude calculus that expresses magnitudes by means of numeric intervals [16], to express (vague) semi-quantitative changes in probability to represent and propagate semi-quantitative expressions of probability. Argumentation schemes 15 and 16 specify argumentation schemes supporting influences with magnitudes expressed by finite differences, which correspond to the absolute order-of-magnitude calculus employed by Parsons. More specifically, argumentation scheme 15 defines context-free order-of-magnitude influences and argumentation scheme 16 context-specific ones.

Argumentation scheme 15: *Argument from Context-Free Order-of-Magnitude Influence*

- B has an effect on the likelihood of C of magnitude m
- Magnitude m is defined by a finite difference between l and u

---

$\therefore$ $l \leq P(c|b,\mathbf{x}) - P(c|\overline{b},\mathbf{x}) \leq u$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Argumentation scheme 16: *Argument for Context-Specific Order-of-Magnitude Influence*

- B has an effect on the likelihood of C of magnitude m, given an assignment b_c of the variables in B_c
- Magnitude m is defined by a finite difference between l and u

---

$\therefore$ $l \leq P(c|b,\mathbf{b_c},\mathbf{x}) - P(c|\overline{b},\mathbf{b_c},\mathbf{x}) \leq u$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

In these schemes, l and u correspond to the lower and upper bounds on the magnitude of the influence. As these order-of-magnitude influences are a generalisation of qualitative influences that allow a more precise definition of an influences effect, these boundary conditions identify the type of qualitative influence this order-of-magnitude influence corresponds to. If $l < u \leq 0$, the argument's antecedent implies a negative qualitative influence and if $0 \leq l < u$, it implies a positive qualitative influence.

Arguments for order-of-magnitude influences provide a means to quantify an expert's beliefs with greater precision. But as such beliefs can take many different forms and be expressed in different ways, a variety of schemes can be proposed. Consider, for instance, the order-of-magnitude calculus NAPIER, which employs rounded logarithms to express the magnitude of a number [38]. In the spirit of this approach, the magnitude of an influence can be expressed by means of a lower and upper bound on the factor by which a probability increases or decreases. Thus, an argument for an order-of-magnitude influence can also be specified as in scheme 17. Here, $0 \leq l < u \leq 1$ corresponds to a negative qualitative influence and $1 \leq l < u$ to a positive qualitative influence.

Argumentation scheme 17: *Argument for Context-Specific Order-of-Magnitude Influence*

- B has an effect on the likelihood of C of magnitude m, given an assignment $\mathbf{b_c}$ of the variables in B_c
- Magnitude m is defined by a factor between l and u

$\therefore$ $l \leq \frac{P(c|b,\mathbf{b_c},\mathbf{x})}{P(c|\bar{b},\mathbf{b_c},\mathbf{x})} \leq u$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 15,16,17

1. If the magnitude is expressed as $l < u \leq 0$ using scheme 15 or 16, or as $0 \leq l < u \leq 1$ using scheme 17, does B have a non-positive effect on C? If the magnitude is expressed as $0 \leq l < u$ using scheme 15 or 16, or as $1 \leq l < u$ using scheme 17, does B have a non-negative effect on C? (tested by the critical questions of Argumentation scheme 5)
2. Is there evidence to support the specification of the magnitude of an influence?
   - Is the claim supported by evidence?
   - Does this evidence concern a magnitude of effect?
   - Does the claim stipulate the boundary conditions of the claim?
3. Is there a set of circumstances $\mathbf{x}$ that can be described by $\mathbf{X}$ in which the magnitude of the effect is not in the range m given $\mathbf{b_c}$: i.e. where the effect is:
   (a) greater than m (i.e. below l in the case of a non-positive effect or above u in case of a non-negative effect), or
   (b) smaller than m (i.e. above u in the case of a non-positive effect or below l in case of a non-negative effect)?
4. Is there a necessary condition for the magnitude of the influence of B on C to be m that is not included in $\mathbf{b_c}$?

For example, an expert in trace evidence may produce an argument on the basis of his/her understanding of the nature of the trace material involved in the case, the events that occurred in the alleged crime and data on random trace transfers and subsequent shedding and data concerning the whereabouts of suspect and victim (i.e. an argument from expert judgement as defined in scheme 4), that trace transfer is a great deal more likely under the guilt hypothesis (g) than it would be if the suspect is a random person unrelated to the crime ($\overline{g}$). The expert may feel confident enough to quantify what he/she means by "a great deal more likely" and suggest that to say "10 times more likely" would be a conservative estimate. In such a situation, scheme 17 can be applied to express the views of the expert.

5.3.3 Arguments for Second Order Influences

As shown in earlier work, context-specific order-of-magnitude influences can be employed to express qualitatively significant "second order influences" [33]. These are relationships between a set of variable assignments and a first order influence, affecting the presence or magnitude of effect of the latter. A disabler, defined by means of argumentation scheme 18 describes conditions in which a first order effect does not hold. In other words, the negation of a disabler is a necessary condition for the first order influence to have effect. An enabler, described by means of argumentation scheme 19 describes conditions that must hold for a first order influence to have effect. In other words, the enabler is a necessary condition for the first order influence to have effect.

Argumentation scheme 18: *Argument for Disabler*

- B can have an effect on the likelihood of C
- $\mathbf{b_c}$ is a disabler of the effect of B on the likelihood of C

$\therefore$ $P(c|b,\mathbf{b_c},\mathbf{x}) = P(c|\overline{b},\mathbf{b_c},\mathbf{x})$ for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 18

1. Can $\mathbf{b_c}$ not be said to have a direct effect on C rather than a second order effect?
2. Is the effect of $\mathbf{b_c}$ to block the influence of B on C or merely to inhibit it?
3. Does the disabling effect assume any circumstances that are explained by variables in X rather than $\mathbf{b_c}$?

Argumentation scheme 19: *Argument for Enabler*

- B can have an effect on the likelihood of C
- $\mathbf{b_c}$ is an enabler of the effect of B on the likelihood of C

$\therefore$ $P(c|b,\overline{\mathbf{b_c}},\mathbf{x}) = P(c|\overline{b},\overline{\mathbf{b_c}},\mathbf{x})$ for any assignment $\overline{\mathbf{b_c}}$ of the variables of $\mathbf{B_c}$ that satisfies $\neg\mathbf{b_c}$ and any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 19

1. Can $\mathbf{b_c}$ not be said to have a direct effect on C rather than a second order effect?
2. Is there no influence of B on C in the absence of $\mathbf{b_c}$, or is the effect of $\mathbf{b_c}$ merely to amplify an influence?

3. Does the enabling effect assume any circumstances that are explained by variables in X rather than $\mathbf{b_c}$?

The cross-transfer model does not contain a good example of an enabler/disabler. It can be argued that contact between victim and suspect (c_t) is an enabler to the effect of the guilt hypothesis (G) on transfer of trace material from suspect to victim (T_v). Specifically, the guilt hypothesis makes transfer of trace material from suspect to victim more likely, but only if there was contact between victim and suspect. However, that line of reasoning is awkward. c_t is really just a necessary condition for t_v rather than a necessary condition on the effect of G on T_v.

An enabler or disabler specifies a necessary condition on an influence between variables and need not make the consequent proposition impossible under certain conditions. In an extensions of the cross-transfer model that specialises in specific crimes, it may be possible to include an additional variable W that differentiates the ways in which a crime could have been perpetrated. Some ways in which the crime could have been perpetrated might justify an increased likelihood of trace transfer and some might not. Here, there are no values of W that constitute a necessary condition on t_v, but under certain values of W, the value of G the probability of t_v would not be affected.

Enablers and disablers are qualitative second order influences on order-of-magnitude first order influences. It is also possible to define second order influences that impact on the magnitude of a first order effect by means of argumentation scheme 20. An inhibitor describes conditions in which a first order influence has a reduced effect compared to conditions in which the inhibitor does not hold. An amplifier describes conditions in which a first order influence has a stronger effect compared to conditions in which the amplifier does not hold.

Argumentation scheme 20: *Argument for* $[Inhibitor|Amplifier]$

- B can have an effect on the likelihood of C
- $\mathbf{b_c}$ is an inhibitor of the effect of B on the likelihood of C with magnitude m
- Magnitude m is defined by a magnitude differential between l and u (with $[l < u < 0|0 < l < u]$)

$\therefore$ $l \leq \left|P(c|b,\mathbf{b_c},\mathbf{x}) - P(c|\overline{b},\mathbf{b_c},\mathbf{x})\right| - \left|P(c|b,\overline{\mathbf{b_c}},\mathbf{x}) - P(c|\overline{b},\overline{\mathbf{b_c}},\mathbf{x})\right|$ $\leq u$ for any assignment $\overline{\mathbf{b_c}}$ of the variables of $\mathbf{B_c}$ that satisfies $\neg\mathbf{b_c}$ and for any assignment $\mathbf{x}$ of the variables in $\mathbf{X}$

Critical Questions for argumentation scheme 20

1. Can $\mathbf{b_c}$ not be said to have a direct effect on C rather than a second order effect?
2. Is there an influence of B on C in the [presence|absence] of $\mathbf{b_c}$?
3. Is there evidence to support the magnitude assessment m?
4. Can the effect of the amplifier/inhibiter be described as a magnitude differential, or does the information in support of it entail another notion of inhibitor/amplifier?

5. Is there a set of circumstances $\mathbf{x}$ that can be described by $\mathbf{X}$ in which the magnitude differential is in the range m?

The cross-transfer model does not contain a good example of an inhibitor or amplifier. Consider an extension of this model that is concerned with cross-transfer of solid trace materials, such as hairs or fibres. In such a model, the time and movements of victim and suspect would become important as trace materials such as hairs and fibres tend to be shed. The total number of traces shed tends to increase over time and with greater movements. In our model and its extension, transfer of trace materials, say, from suspect to victim makes retrieval of trace material of the suspect from the victim's body more likely. Shedding, as affected by time and movement, acts as an inhibitor on that relationship between transfer and retrieval. Because shedding reduces the number of transferred traces that remain on the victim's body, it becomes less likely that any transferred traces can still be found.

6 Case Study

This section examines how the approach presented herein is applied by means of a practical case study. It demonstrates how the argumentation schemes defined above are employed to construct arguments for constraints on conditional probability tables and likelihood ratio assessments on the value of evidence, and how the critical questions associated with these arguments and analysis approaches facilitate validation of probabilistic models.

The case study employed herein is Aitken, Taroni and Garbolino's cross-transfer model [3] because its structure provides a solution to a real-world model, it possesses the complexity of real-world evidential reasoning problems and its conditional probability tables include subjective probabilities, probabilities that can be extracted from data sets and probabilities that stem from the model structure and variable definition.

6.1 Hybrid Argument-Probabilistic Model of the Value of Evidence

The structure of Aitken, Taroni and Garbolino's cross-transfer model is introduced in Section 3.1.2 and is presented in Figure 1. In the original paper, this structure is carefully justified. A set of conditional probability tables are proposed. However, instead of justifying the individual values in these conditional probability tables, values that are potentially subjective in nature are identified and a sensitivity analysis is performed over these values. The cross-transfer model can be applied to a broad range of cases. In these case studies, a scenario is considered where the alleged crime g concerns violent altercation that escalates into an assault leading to the death of the victim, and the evidence concerns transferred blood traces as a result of punches being exchanged.

The approach taken in this paper is to construct arguments implying constraints on the values in the conditional probability tables, justified by ancillary evidence and instantiated from the argumentation schemes introduced in Sections 4 and 5. Due to the way in which Bayesian networks define conditional independence relationships, the probability distribution of each variable only needs to be considered in relation to its immediate parents in the DAG, thereby decomposing the model construction problem into a range of smaller ones. In the remainder of this subsection, each of these smaller subproblems is considered in turn. The arguments instantiated from the argumentation schemes will be presented in the form of simple argument diagrams from ancillary evidence to constraint. Because the original paper did not take this approach, this section contains a reverse engineered set of arguments in support of conditional probability distributions that were proposed. While this exercise is based on the discussion in the paper and informal discussions with the authors, these arguments do not necessarily correspond to the views of the authors of the model. The reader is reminded that what is important for the purposes of this paper is that a qualitative justification for conditional probability distributions can be formulated and validated, and that the implications of rejecting a justification can be computed.

No prior probability distribution for G, representing the guilt/innocence hypothesis will be constructed. As the probability values to be computed as part of the likelihood ratio are conditioned on defence and prosecution hypotheses that specify whether the suspect is guilty or innocent, this need not be specified. Background information concerning the suspect and victim with regards to the potential presence of trace material may or may not be available. If it is not available at the start of an investigation or court case, it may become available. For that reason, arguments constraining the prior probability distributions of B_v and B_s will be constructed, but these may be overridden as new direct evidence becomes available.

Figure 3 is an argument diagram concerning the conditional probability table of C_t, representing contact between victim and suspect. C_t has only one parent variable: G. The argument diagram contains two sets of arguments, resulting in one constraint each. The first argues that scenarios where the suspect is guilty (g) entails contact. Thus g is a sufficient condition for c_t. The second argues that, given the circumstances of the case, contact between victim and suspect is not very likely in circumstances other than the guilt scenario (though not impossible). Specifically, the improbability is defined by an upper bound of 1% on the probability of c_t in the absence of probable explanation. Figure 3 shows the constraints these arguments imply and the conditional probability table admitted by the constraints.

Figure 4 is an argument diagram concerning the prior probability distribution of B_i and the conditional probability table of P_i, with $i = v, s$. It is a post-hoc justification for the sample values presented in the model by Aitken et. al. [3]. In the model, it is argued (from data) that only a small proportion of the population (less than 0.1%) comes into contact with blood spatter (b_i). Then, it is argued (from expert judgement) that if there is no reason to believe that the probability is higher in this particular case, and an upper bound on the probability distribution of b_i is known, then that upper bound is an upper bound on $P(b_i)$.

In this approach to modelling the probability distribution of P_i, the authors equate the background predisposition of individual i to the presence of trace material with the presence of trace material on i. In other words, here, P_i is defined as B_i. This is a sensible approach if B_i can only be modelled as P_i, as is the case when no background attributes of i imply a particular predisposition to the presence of trace material.

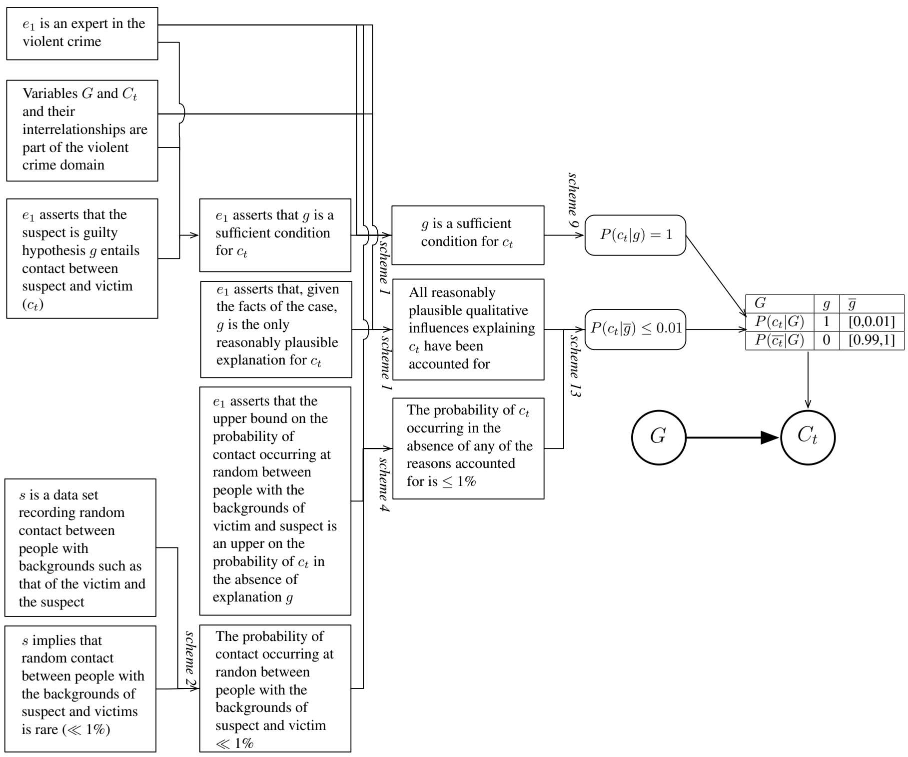
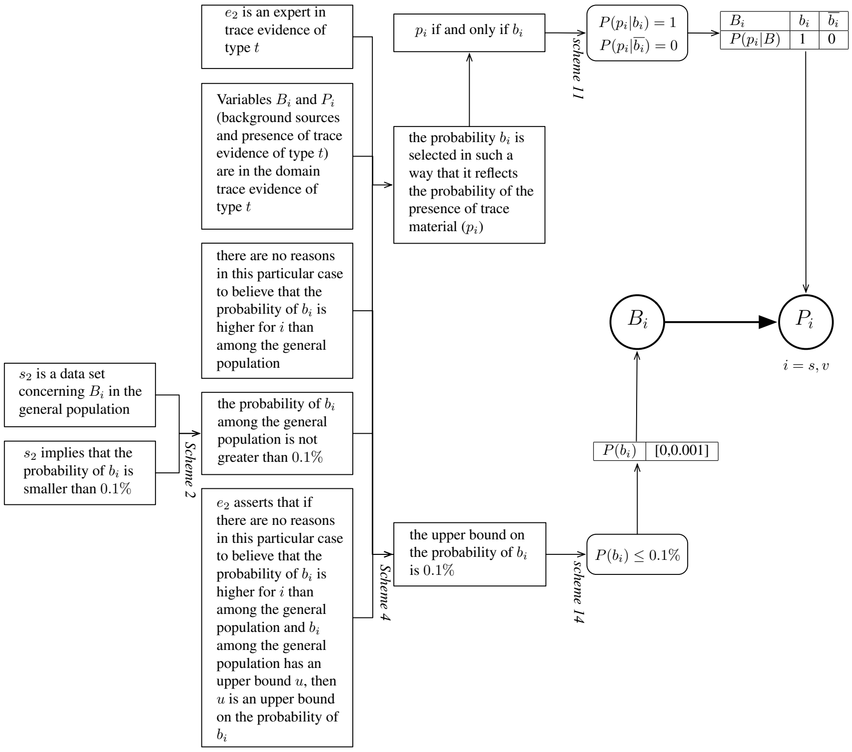


Fig. 3 Argument Diagram of Arguments Supporting the Conditional Probability Table of C_t

These arguments should be adapted in certain other circumstances. Firstly, evidence may be available that the background of the victim or suspect is such that (s)he has come into contact with blood spatter prior to the crime under consideration: i.e. it is known that b_v or b_s is true. Secondly, even if is not known that b_v or b_s is not true, there may be evidence to suggest that there are reasons for $P(b_i)$ to be greater than among the general population, e.g. because of i's profession, behaviour or medical condition. Thirdly, B_i does not necessarily define P_i, particularly when there are certain reasons to believe $P(b_i)$ to be greater for i than among the general population.

Figure 5 is an argument diagram concerning (potential) evidence of trace transfer from victim to suspect and from suspect to victim. It presents arguments constraining the conditional probability tables of two types of variable. The specification of both conditional probability tables is very similar. Variable M_i identifies whether a test comparing traces yields a match. As such, M_i is a classification of C_i. Variable RM_i identifies whether a match is reported. As such, RM_i is the observation of M_i. Both types of argument are instances of scheme 12. The specific values in Figure 5 stem from Aitken, Taroni and Garbolino's paper, which suggest that the type I error of classifying C_i (in M_i) is not greater than 0.01%, the type I error of observing M_i (as RM_i) is not greater than 0.1% and the type II error of both is 0.



B_i	b_i	$\overline{b_i}$
$P(p_i\|B)$	1	0

Fig. 4 Argument Diagram of Arguments Supporting the Conditional Probability Tables of B_v, P_v, B_s and P_s

Figure 6 is an argument diagram concerning the conditional probability table T_v. This figure illustrates how a model can be constructed that contains arguments for qualitative features, extreme probabilities and quantitative features. A similar model could be constructed for T_s. In general, it could be argued, however, that the probability of blood spatter transfer from victim to suspect (t_s) is somewhat greater than the probability of blood spatter from suspect to victim (t_v) because the victim is more likely to be injured more severely. To avoid overcomplicating the example, this will be ignored herein and the more conservative model of Figure 6 will be applied to T_s as well.

Finally, Figure 7 is an argument diagram concerning the conditional probability tables of variables C_v and C_s. This diagram is constructed along the same principles as the preceding ones.

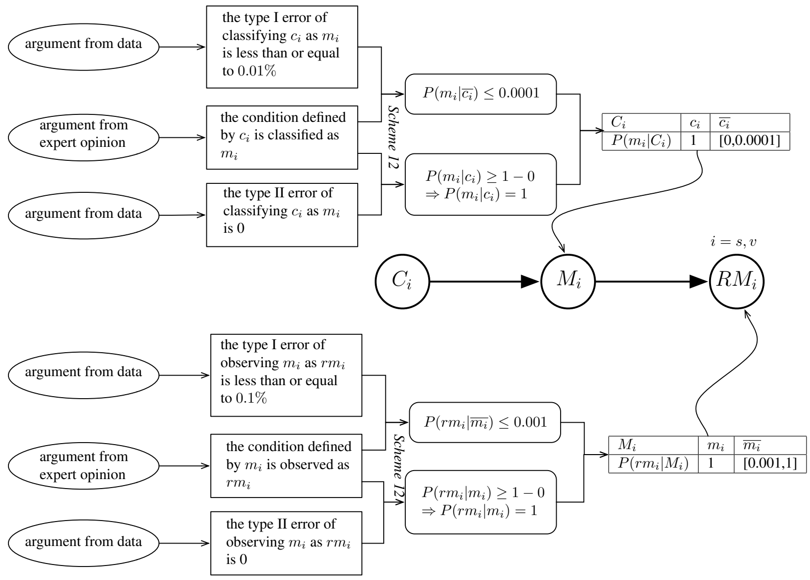


Fig. 5 Argument Diagram of Arguments Supporting the Conditional Probability Tables of C_v, M_v, RM_v, C_s, M_s and RM_s

6.2 Analysis and Validation

Once the argument models have been specified, their implications can be computed and the arguments validated. This section demonstrates this by means of the argumentation models of the previous subsection.

For the purposes of computing the implications of argumentation models, two algorithms are needed that are beyond the scope of this paper. First, the set of solutions to the constraint problems that follow from the argumentation models need to be computed. A conventional constraint satisfaction algorithm was used, though a range of algorithms can be employed for this purpose as described in 3.2.2. In Figures 3–7, the solution space implied by the constraints supported by arguments in the model is depicted. Second, the extreme points of the solutions space defined in this way are used to define convex sets of conditional probability values, which are propagated in a Bayesian network. This is accomplished with a credal network tool [13]. In this exercise, JavaBayes was employed [14].

Let **A** be the set of arguments presented in the previous subsection. The value of the evidence based on **A** is computed by (4). In the ongoing example, consider, say, two pieces of evidence: blood spatter matching the suspect is retrieved from the victim ($e_v = rm_v$) and blood spatter retrieved from the suspect does not match the victim ($e_s = \overline{rm_s}$). Also consider two hypotheses: a prosecution hypothesis that the suspect is guilty of the crime ($h_p = \{g\}$) and a defence hypothesis that the suspect is not guilty of the crime ($h_d = \{\overline{g}\}$). Note that these hypotheses do not include assumptions about

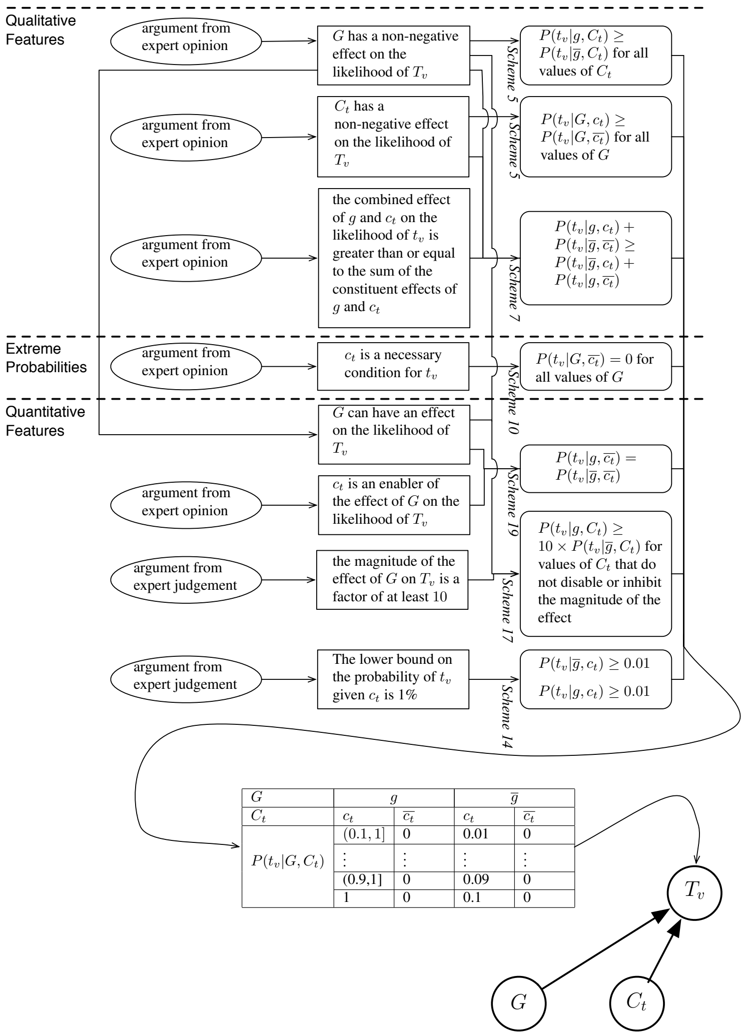


Fig. 6 Argument Diagram of Arguments Supporting the Conditional Probability Tables of G, C_t and T_v

the backgrounds of the suspect and victim. In this case, the value of the evidence is:

$$LR_{\mathbf{A}}(e_v|h_p,h_d) = \frac{P_{\mathbf{A}}(RM_v|g)}{P_{\mathbf{A}}(RM_v|\overline{g})} = [83,\infty)$$

$$LR_{\mathbf{A}}(e_s|h_p,h_d) = \frac{P_{\mathbf{A}}(RM_s|g)}{P_{\mathbf{A}}(RM_s|\overline{g})} = (0,0.901]$$

This implies that e_v provides some support for h_p over h_d and that e_s provides some support for h_d over h_p. The meaning of the likelihood ratio values in the intervals are interpreted in the same way as those computed by conventional Bayesian techniques

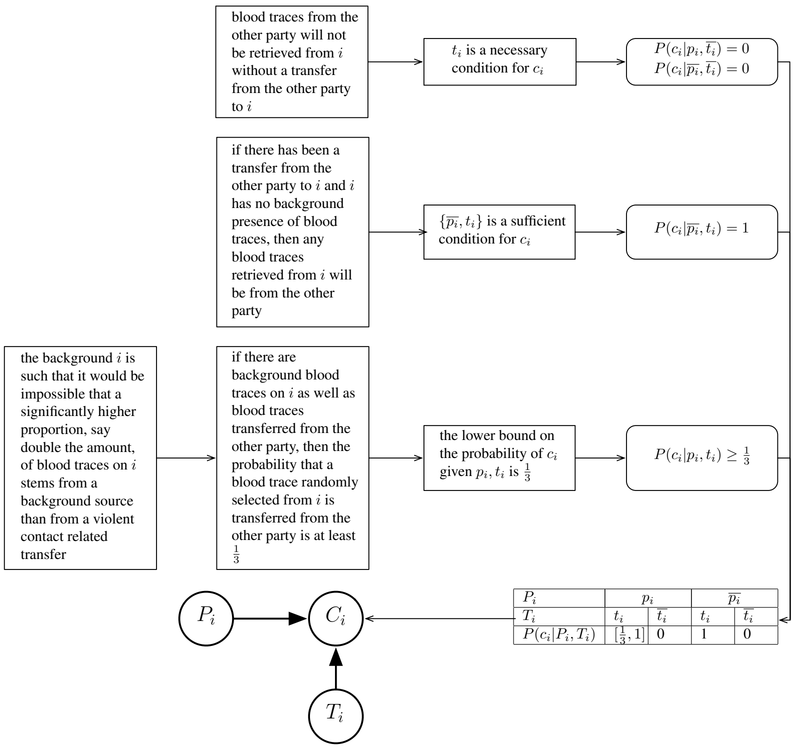


Fig. 7 Argument Diagram of Arguments Supporting the Conditional Probability Tables of C_v and C_s

because they have the same meaning. But, the arguments in **A** are not sufficiently precise to differentiate between the different values in the range. In other words, the approach presented herein computes a conservative assessment of the value of evidence based on sets of likelihood ratios supported by arguments. Using the interpretation proposed by Cook et. al., these ranges of likelihood ratios suggest that e_v provides at least moderate support for the prosecution hypothesis over the defence hypothesis and that e_s provides at least very weak support for the defence hypothesis over the prosecution hypothesis.

The value of evidence assessments depend on the acceptance of the arguments in **A**. These arguments are validated by applying the critical questions associated with the argumentation schemes that the arguments are instantiated from, and using these critical questions as a means to find reasons to reject the argument. However, by computing the effect of rejecting a particular argument, this validation effort can be focussed on those arguments with the most significant impact. The impact of an argument is computed by (5).

Consider, for instance, the argument model presented in Figure 7 with respect to C_v, which implies that i is instantiated to v in this model. This model contains an assertion that the background of the victim is such that it would be impossible that

P_i	p_i		$\overline{p_i}$	
T_i	t_i	$\overline{t_i}$	t_i	$\overline{t_i}$
$P(c_i\|P_i,T_i)$	$[0,1]$	0	1	0

Table 2 Effect of removing argument a

G	g		$\overline{g}$	
C_t	c_t	$\overline{c_t}$	c_t	$\overline{c_t}$
$P(t_v\|G,C_t)$	(0.5,1]	0	0.05	0
	⋮	⋮	⋮	⋮
	(0.9,1]	0	0.09	0
	1	0	0.1	0

Table 3 Effect of adding argument a'

a significantly higher proportion, say double the amount of blood traces on v stems from a background source other than s. This implies that $P(c_v|p_v,t_v) \geq \frac{1}{3}$. In considering the second of the critical questions of argumentation scheme 14, it is clear that the assertion involves a comparison based on the relative occurrence of two events: the transfer of individual units of blood spatter to the victim in the course of the victim's background activities and the transfer of individual units of blood spatter from suspect to victim in the crime under investigation. For the purpose of demonstrating use of calculating the effect of an argument on the outcome of analysis, suppose that it is argued that the aforementioned to types of transfer are difficult to compare with one another (i.e. the answer to critical question 3a is "no"). Indeed, the argumentation model presented in Figure 7 does not include evidence of absolute frequencies and it may not be deemed appraise to specify the above boundary condition.

For these reasons, the entire argument may be deemed suspect. If this line of reasoning is deemed to defeat the assertion, then the associated argument a, including constraint $P(c_v|p_v,t_v) \geq \frac{1}{3}$ is removed from the problem. This affects the conditional probability table in Figure 7 as shown in Table 2. The effect of removing argument a on the value of evidence e_v can be computed by means of (5) as:

$$LR_{\mathbf{A}/\{\mathbf{a}\}}(e_v|h_p,h_d) = \frac{P_{\mathbf{A}/\{a\}}(RM_v|g)}{P_{\mathbf{A}/\{a\}}(RM_v|\overline{g})} = [83,\infty]$$

In other words, the value of the evidence is unaffected by this alteration (there is some effect if more significant digits would be presented, but the effect is negligible). As such, defeating this argument that the earlier analysis depended on does not invalidate its conclusion.

If a more precise assessment of the value of evidence is required, arguments must be formulated that further constrain the conditional probability tables. However, not all new arguments are equally informative. Yet, all new arguments ought to be validated. By means of (6), the impact of a new argument can be computed.

For instance, consider a hypothetical situation where the suspect had a tooth removed recently and some data that suggests that in 5% of all cases of contact with such a person, there could be some transfer of blood spatter from a person who had a tooth removed recently. One might argue, rightly or wrongly, that that implies that there is a lower bound on the probability of transfer of blood spatter from suspect to victim given that there was contact between suspect and victim, irrespective of whether the suspect is guilty of the crime. But, any debate about this ancillary evidence and its implications is only meaningful if the additional argument a' has a significant impact on the assessment of the value of the evidence. To evaluate the impact of this prospective new argument, new constraints $P(t_v|g,c_t) \geq 0.05$ and $P(t_v|\overline{g},c_t) \geq 0.05$ are added to the argument model of Figure 6. This additional constraint reduces the solution space to the conditional probability table of T_v as specified in Table 3. The effect of the additional argument a' can be computed by means of (6) as:

$$LR_{\mathbf{A}\cup\{\mathbf{a}'\}}(e_v|h_p,h_d) = [416,\infty]$$

In other words, the introduction of a' increases the value of evidence e_v with at least a factor 5. As such, the impact of such an argument would be significant and therefore, merits consideration.

The above forms of analysis achieve the same objectives that the sensitivity analysis of an expert formulating a Bayesian model aims to achieve. However, the key difference of the approach presented herein is that it reasons with concrete, qualitatively significant arguments rather than abstract numbers whose meaning and implications can be hard to understand. As such, this work presents a significant step towards the integration of argumentation and Bayesian based approaches. This allows it to provide some transparency in the way probabilistic models and methods of analysis lead to conclusions, to help identify how strongly the results depend on subjective judgement and to support the validation of such subjective judgement.

7 Conclusions and Future Work

This paper has presented a novel approach to enable argumentation about probability distributions in probabilistic models for evidential reasoning in law. It has introduced a set of argumentation schemes for the construction of argumentation models concerning probability distributions and associated sets of critical questions for their validation. The use of the proposed approach has been demonstrated by means of an extended case study concerning a complex realistic probabilistic model for evidential reasoning.

In this way, this paper addresses an important concern of the use of probabilistic models for evidential reasoning in law: the use of subjective probability in assessing the value of evidence. Existing approaches for reasoning with subjective probability allow for subjective probabilities to be represented as numbers, numerical intervals or sets of numerical intervals that express a person's belief in certain propositions, possibly under certain conditions. But with these approaches, there are no means to understand where these numbers come from and what justifies, according to the modeller, their use as an expression of belief. This is an important drawback in legal applications where inaccurate assessments of the value of evidence can have serious implications, such as miscarriages of justice.

In the development of this work, a number of issues have been identified that need to be addressed in future work. Firstly, the nature of the argumentation models proposed herein provide opportunities for producing more informed explanations of the outcome of probabilistic analysis. In earlier work, the author has developed a technique for the extraction of argumentation models from Bayesian networks. This could be extended to incorporate the argumentation models considered in this paper.

Secondly, the results of probabilistic models are not only affected by numerical values, but also by structural ones: the choice of variables to include in the model and the conditional or unconditional independence relationships that are assumed between variables. Again, while a range of modelling techniques for these features exist, such as Bayesian networks, these do not allow for the construction of models of their rationale with a view to facilitate their validation.

Thirdly, as explained in Section 3, the computation of the implications of accepting arguments concerning conditional probability distributions and changes in the set of accepted arguments involves computing solutions of a dynamic continuous constraint satisfaction problem. This type of problem has not been studied extensively, even though a wide range of solution algorithms exist for constraint problems that are either dynamic or continuous, but not both. Future work aims to explore the integration of such algorithms.

Acknowledgement

I would like to thank the anonymous reviewers for their very constructive, in-depth reviews of an earlier version of this article.

References

1. Adams, E.: The Logic of Conditionals. Reidel (1975)
2. Aitken, C., Taroni, F.: Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd edition edn. Wiley (2004)
3. Aitken, C., Taroni, F., Garbolino, P.: A graphical model for the evaluation of cross-transfer evidence in DNA profiles. Theoretical Population Biology **63**, 179–190 (2003)
4. of Appeal (Criminal Division), C.: Regina v T, redacted judgement approved by the court for handing down (subject to editorial corrections). EWCA Crim 2439 (2010)
5. of Appeal (Queen's Bench Division), C.: R v Turnbull. QB 224 (1977)
6. Berger, C., Buckleton, J., Champod, C., Evett, I., Jackson, G.: Evidence evaluation: A response to the appeal court judgment R v T. Science and Justice **51**, 43–49 (2011)
7. Bex, F., van den Braak, S., van Oostendorp, H., Prakken, H., Verheij, B., Vreeswijk, G.: Sense-making software for crime investigation: how to combine stories and arguments. Law, Probability and Risk **6**, 145–168 (2007)
8. Bex, F., Verheij, B.: Solving a murder case by asking critical questions: An approach to fact-finding in terms of argumentation and story schemes. Argumentation **26**, 325–353 (2012)
9. Biedermann, A., Taroni, F., Delemont, O., Semadeni, C., Davison, A.: The evaluation of evidence in the forensic investigation of fire incidents. part ii. practical examples of the use of Bayesian networks. Forensic Science International **147**, 59–69 (2005)
10. Bowen, R., Schneider, J.: Forensic databases: Paint, shoe prints, and beyond. National Institute of Justice Journal **258**, 34–38 (2007)
11. Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J.: A model for case assessment and interpretation. Science and Justice **38**(6), 151–156 (1998)

12. Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J.: A model for case assessment and interpretation. Science and Justice **38**(6), 151–156 (1998)
13. Cozman, F.: Credal networks. Artificial Intelligence **120**, 199–233 (2000)
14. Cozman, F.: JavaBayes Bayesian networks in java (2013). URL http://www.cs.cmu.edu/ javaBayes/
15. CPS: The code for crown prosecutors. https://www.cps.gov.uk/publications/docs/code_2013_accessible_english.pdf
16. Dague, P.: Numeric reasoning with relative orders of magnitude. In: In Proceedings of the 11th National Conference on Artificial Intelligence, pp. 541–547 (1993)
17. Dague, P.: Symbolic reasoning with relative orders of magnitude. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 1509–1514 (1993)
18. Dawid, P.: Probability and statistics in the law. In: Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics, pp. 89–95 (2005)
19. Dechter, R.: Constraint Processing. Morgan-Kaufmann (2003)
20. Druzdzel, M., van der Gaag, L.: Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In: Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence, pp. 141–148 (1995)
21. Druzdzel, M., van der Gaag, L.: Building probabilistic networks: where do the numbers come from? IEEE Transactions on Knowledge and Data Engineering **12**(4), 481–486 (2000)
22. Evett, I., Lambert, J., Buckleton, J.: A Bayesian approach to interpreting footwear marks in forensic casework. Science and Justice **38**(4), 241–247 (1998)
23. Fenton, N.: Science and law: Improve statistics in court. Nature **479**, 36–37 (2011)
24. Fenton, N., Berger, D., Lagnado, D., Neil, M., Hsu, A.: When 'neutral' evidence still has probative value (with implications from the Barry George case) (2013). Http://www.scienceandjusticejournal.com/article/S1355-0306(13)00059-2
25. Fenton, N., Lagnado, D., Neil, M.: A general structure for legal arguments using Bayesian networks. Cognitive Science **37**(61–102) (2013)
26. Fenton, N., Neil, M.: On limiting the use of Bayes in presenting forensic evidence. https://www.eecs.qmul.ac.uk/Downloaded 15/1/2012 (2012)
27. Garbolino, P., Taroni, F.: Evaluation of scientific evidence using Bayesian networks. Forensic Science International **125**, 149–155 (2002)
28. Gent, I., Jefferson, C., Miguel, I.: MINION: A fast, scalable, constraint solver. In: Proceedings of the 17th European Conference on Artificial Intelligence, pp. 98–102 (2006)
29. Gittelson, S., Biedermann, A., Bozza, S., Taroni, F.: Modeling the forensic two-trace problem with Bayesian networks. Artificial Intelligence and Law **21**(2), 221–252 (2013)
30. Goldszmidt, M., Pearl, J.: Qualitative probabilities for default reasoning, belief revision, and causal modeling. Artificial Intelligence **84**, 57–112 (1996)
31. Granvilliers, L., Benhamou, F.: Realpaver: An interval solver using constraint satisfaction techniques. ACM Transactions on Mathematical Software **32**, 138–156 (2006)
32. Hepler, A., Dawid, P., Leucari, V.: Object-oriented graphical representations of complex patterns of evidence. Law, Probability and Risk **6**(1–4), 275–293 (2007)
33. Keppens, J.: Towards qualitative approaches to Bayesian evidential reasoning. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law, pp. 17–25 (2007)
34. Keppens, J.: Conceptions of vagueness in subjective probability for evidential reasoning. In: Proceedings of the 22nd Annual Conference on Legal Knowledge and Information Systems, pp. 89–99 (2009)
35. Keppens, J.: Argument diagram extraction from evidential Bayesian networks. Artificial Intelligence and Law **20**(2), 109–143 (2012)
36. Keppens, J., Shen, Q., Price, C.: Compositional Bayesian modelling for computation of evidence collection strategies. Applied Intelligence **35**(1), 134–161 (2011)
37. Mortera, J., Dawid, A., Lauritzen, S.: Probabilistic expert systems for DNA mixture profiling. Theoretical Population Biology **63**, 191–205 (2003)
38. Nayak, P.: Order of magnitude reasoning using logarithms. In: Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning, pp. 201–210 (1992)
39. Niederliński, A.: A Quick and Gentle Guide to Constraint Logic Programming ECLiPSe. PKJS (2011)
40. Nordgaard, A., Rasmusson, B.: The likelihood ratio as value of evidence–more than a question of numbers. Law, Probability and Risk **11**(4), 303–315 (2012)
41. Parsons, S.: Qualitative probability and order of magnitude reasoning. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **11**(3), 373–390 (2003)
42. Parsons, S.: On precise and correct qualitative probabilistic inference. International Journal of Approximate Reasoning **35**(2), 111–135 (2004)

43. Parsons, S., Mamdani, E.: On reasoning in networks with qualitative uncertainty. In: Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (1993)
44. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, second edition edn. Morgan-Kaufmann (1988)
45. Prakken, H.: Analysing reasoning about evidence with formal models of argumentation. Law, Probability and Risk **3**(1), 33–50 (2004)
46. Raiman, O.: Order of magnitude reasoning. Artificial Intelligence **51**, 11–38 (1991)
47. Redmayne, M., Roberts, P., Aitken, C., Jackson, G.: Forensic science evidence in question. Criminal Law Review **5**, 347–356 (2011)
48. Reiter, R.: On closed world data bases. In: H. Gallaire, J. Minker (eds.) Logic and Data Bases, pp. 119–140. Plenum Press (1978)
49. Renooij, S., van der Gaag, L., Parsons, S.: Context-specific sign-propagation in qualitative probabilistic networks. Artificial Intelligence **140**, 207–230 (2002)
50. Robertson, B., Vignaux, G., Berger, C.: Extending the confusion about Bayes. Modern Law Review **74**, 444–455 (2011)
51. Sam-Haroud, D., Faltings, B.: Consistency techniques for continuous constraints. Constraints **1**(1), 85–118 (1996)
52. Schum, D.: The Evidential Foundations of Probabilistic Reasoning. Northwestern University Press (1994)
53. Spohn, W.: A survey of ranking theory. In: F. Huber, C. Schmidt-Petri (eds.) Degrees of Belief, *Synthese Library*, vol. 342, pp. 185–228. Springer Science + Business Media (2009)
54. Thompson, W.: Hard cases make bad law – reactions to R v T. Law, Probability and Risk **11**(4), 347–359 (2012)
55. Tsang, E.: Foundations of Constraint Satisfaction. Academic Press (1993)
56. Verfaillie, G., Schiex, T.: Solution reuse in dynamic constraint satisfaction problems. In: Proceedings of the 12th National Conference on Artificial Intelligence, pp. 307–312 (1994)
57. Vlek, C., Prakken, H., Verheij, B.: Modeling crime scenarios in a Bayesian network. In: In Proceedings of the 14th International Conference on Artificial Intelligence and Law, pp. 150–159 (2013)
58. Walton, D.: Appeal to Expert Opinion. Penn State University Press (1997)
59. Walton, D., Godden, D.: The nature and status of critical questions in argument schemes. In: D. Hitchcock (ed.) The Uses of Argument, pp. 476–484. McMaster University (2005)
60. Walton, D., Reed, C., Macagno, F.: Argument Schemes. Cambridge University Press (2008)
61. Weichselberger, K.: The theory of interval-probability as a unifying concept for uncertainty. International Journal of Approximate Reasoning **24**, 149–170 (2000)
62. Wellman, M., Henrion, M.: Explaining ”explaining away”. IEEE Transactions on Pattern Analysis and Machine Intelligence **15**, 287–291 (1993)
63. Wild, C., Seber, G.: Chance Encounters: A First Course in Data Analysis and Inference. John Wiley & Sons (1999)